UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA VEGETAL

Study and Characterization of CFTR Mutations *in Vitro* and in Native Tissues from non CF Patients with Chronic Airways Diseases

Arsénia Joana Massinga

Mestrado em Biologia Molecular e Genética

Dissertação orientada por:
Professora Doutora Margarida D. Amaral

2017

Table of Contents

List of Figures

List of tables

Acknowledgments/Agradecimentos

Ao concluir este trabalho, além do meu agradecimento à Deus, pelo dom da vida, e ao projecto Caribu do programa Erasmus Mundus, por ter financiado os meus estudos em Portugal, não posso deixar de agradecer às pessoas que de forma directa e indirecta contribuíram para a realização deste trabalho.

Em primeiro lugar, quero agradecer à Professora Margarida Amaral, minha supervisora, por me aceitar no seu grupo e me ter recebido de braços abertos desde o primeiro momento. Agradeço também por me ter dado a oportunidade de desenvolver o meu trabalho, pela disponibilidade e confiança depositada.

Meu muito obrigada à Professora Rita Zilhão, minha mãe de coração, sem ela meus estudos em Portugal, meu trabalho, nada disto teria acontecido. Obrigada por ter apostado em mim, por ter movido mundos e fundos para que eu conseguisse fazer o meu mestrado em Portugal. Obrigada pelo apoio incondicional, pelo carinho e amor com que sempre me tratou, e por estar sempre disponível quando mais precisava. Agradeço também por me ter ajudado a resolver todos os problemas que tive durante a realização das experiências, principalmente nas clonagens e tranfecções.

Agradeço aos meus pais, que mesmo distantes sempre fizeram-me sentir amada e apoiaram todas as minhas decisões. Obrigada as minhas irmãs, Esperança e Rosiménia pelo apoio e simplesmente por existirem, sei que posso contar sempre convosco – porque somos as três mosqueteira: *uma por todas e todas por uma*.

Obrigada a todos os meus colegas de laboratório, com os quais passei momentos felizes de festas, jantares e muito trabalho, e claro momentos *stressantes* também, quando as experiências não corriam bem. Todos vocês ajudaram-me a crescer pessoal e intelectualmente. Agradeço a todos pelos puxões de orelha, pelas chamadas de atenção quando fazia algo mal, graças a todos vocês comecei a redobrar o cuidado em tudo e a ser mais atenciosa. E como num grupo existem sempre pessoas especiais, quero agradecer de maneira muito especial à Verónica, que mais do que uma colega de trabalho e vizinha de bancada, tornou-se numa grande amiga e quase irmã, por me ter ensinado tudo o que sei sobre genotipagem de pacientes e clonagens, por estar sempre disponível, mesmo quando ocupada. Agredeço pela paciência, atenção e carinho que sempre teve comigo (saiba que para onde eu for levarei comigo os anjinhos de natal que me deste, que são a representação do que tens sido para mim). Agradeço a Sara Canato e ao João Santos pelas brincadeiras, e por me suportarem nos momentos de loucura, e também pelas conversas filosóficas. Agradeço à Íris Lameiro e ao Nikhil pelo apoio nos momentos de crise e desânimo, quero que saibam que deram-me forças para continuar, muito obrigada pela amizade. Muito obrigada à Susana Igreja pelas correcções do trabalho, por ter se esforçado muito que toda história fizesse sentido, e por não ter desistido de mim. Obrigada ao Simão pelas conversas beatas e pelos conselhos sábios que sempre me deu. Agradeço à Joana Lérias pela amizade e pelos conselhos. E claro agradeço à Sofia Correia, nossa lab manager, pelo esforço em manter a ordem no laboratório, pelo apoio moral e pela amizade incondicional.

Meu muito obrigada aos meus colegas de mestrado, de maneira especial ao Clemente, meu amigo de longa data, por me ajudar a traçar objectivos de longo prazo, pelo apoio incondicional e pelo esforço que faz para que eu melhore a cada dia. Obrigada também à Carolina Sousa que mais que uma colega tornou-se a minha *miga* do estoiro do orçamento, obrigada pela amizade, companheirismo e por estar sempre disposta a ouvir-me, muito obrigada port udo.

E finalmente agradeço às meninas das Lanjeiras, minha primeira residência, minha grande família portuguesa por me ensinarem a organizar melhor os meus estudos e vida no geral, de maneira muito especial à Isabel Coimbra e a Marta Antunes, pela amizade, pelo carinho com que sempre me trataram. Muito obrigada à família Goes e ao padre José de Melo por me terem acolhido quando mais precisei, e foi graças a isso que consegui escrever esta dissertação.

Khanimambo! Ni khensa swinene!

Summary

Cystic Fibrosis (CF) is the most common autosomal recessive disorder among Caucasians caused by mutations in the CF Transmembrane Conductance Regulator (CFTR) gene, which encodes a protein localized in the apical plasma membrane (PM) of epithelial cells that functions as a chloride (Cl^-) and bicarbonate (HCO_3^-) channel.

Previous studies have shown the occurrence of at least one CFTR mutation also in non-CF patients with asthma, chronic bronchiectasis and chronic obstruction pulmonary disease (COPD) with unknown etiopathology.

Until now about 2,000 CFTR variations have been found, in which 11% of them are known to be splicing mutations. A novel antisense oligonucleotide (AON) therapy approach corrects splicing defects and is giving some hope for patients carrying this type of mutations.

The objective of the current MSc project was two-fold, namely:

1) To investigate the possible involvement of the 9 most common Portuguese CFTR mutations in causing non-CF respiratory disorders: asthma, COPD and chronic bronchiectasis.
2) To characterize the impact of four splicing mutations located in the same splicing consensus-711+1G>T, 711+3A>T, 711+3A>G and 711+5G>A and correct them using a single AON.

Our results showed that at least 3 mutations out of 9 most common Portuguese CFTR mutations were found in our cohort of patients with respiratory disorders.

Concerning the study of splicing mutations, *in vitro* experiments were performed using a reporter splicing minigene. We showed that all four splicing mutations located in the same splicing consensus one IVS5 caused skipping of exon 5, which produce a smaller protein and lead to a mislocalization of the CFTR protein in the PM. In addition, the same AON potentially corrected all four splicing mutations.

Altogether these data suggest that there is an involvement of some CFTR mutations commonly found in Portugal in causing non-CF chronic airway diseases and that patients carrying the splicing mutations studied here can benefit from the same AON for therapy.

Resumo

A fibrose quística (FQ) é a doença autossómica recessiva mais comum na população Caucasiana, afetando 1 em cada 6000 portugueses. Esta doença afeta vários órgãos e os doentes apresentam variados sintomas, os mais comuns sendo: elevadas concentrações de cloreto no suor, infeções bacterianas recorrentes devido ao muco espesso formado nos pulmões e malnutrição. A FQ é causada por mutações no gene da CFTR (do inglês *Cystic fibrosis transmembrane conductance regulator*) que codifica uma glicoproteína localizada na membrana apical de células epiteliais e funciona como um canal de cloreto (Cl^-) e bicarbonato (HCO_3^-).

Tradicionalmente, o diagnóstico da FQ baseia-se nas manifestações clínicas, história familiar e dois testes de suor positivos (>60mM), porém quando esta abordagem é inconclusiva, são feitos testes genéticos para identificar mutações nos dois alelos e medições de corrente para verificar a função ou disfunção da CFTR.

Estudos prévios reportaram uma parcial disfunção da proteína CFTR em doentes sem nenhuma manifestação clínica de FQ, mas com outras doenças como pancreatite, asma, bronquiectasias crónicas e doenças pulmonares de obstrução crónica (DPCO), de origem desconhecida, sugerindo uma associação entre mutações no gene da CFTR e outras doenças que não sejam FQ. Adicionalmente, foi demonstrado que alguns desses indivíduos possuíam apenas uma mutação no gene da CFTR. Contudo, as mutações encontradas são raras e portanto não incluídas entre as mutações mais frequentes encontradas em Portugal. Além disso nem todas as variantes na CFTR estão descritas como causadoras de FQ.

Atualmente são conhecidas mais de 2.000 variações no gene da CFTR, e apenas aproximadamente 250 mutações estão descritas como sendo causadoras de FQ. Estas mutações são classificadas de acordo com o defeito causado na proteína e estão agrupadas em sete classes (I-VII). As mutações de *splicing* pertencem à classe V (mutações que diminuem os níveis da proteína na membrana apical) e compreendem uma fração de 11% das variações conhecidas no gene da CFTR. Estas mutações resultam num *splicing* anormal destruindo as junções normais dos RNAs ou criando um novo local de *splicing* no interior de um intrão ou de um exão. Tais alterações levam à produção de uma proteína disfuncional ou truncada o que diminui os níveis da proteína normal que é expressa.

É previsível que a percentagem de mutações de *splicing* sejam superiores a 11%, devido às mutações *missense*, que não estão totalmente caracterizadas e que também podem levar a um *splicing* anormal. Os doentes que possuem esta classe de mutações podem beneficiar apenas de um fármaco que potencia a função da proteína resultante de transcritos normais, o Ivacaftor (VX770). Porém, como algumas mutações de *splicing* diminuem drasticamente os níveis de proteína funcional, potenciar a proteína não é suficiente, sendo necessário corrigi-la. Para este efeito uma nova abordagem tem sido adotada nomeadamente, o uso de terapias de RNA que corrigem o defeito a nível do RNA mensageiro (mRNA). Uma dessas terapias é o uso de AONs (do inglês *antissense oligonucleotides*), que bloqueiam ou aumentam o processo de *splicing*, de acordo com a sequência a que se ligarão.

No presente trabalho de Mestrado tivemos dois objetivos principais, nomeadamente:

1) Investigar o envolvimento das nove mutações do gene da CFTR mais comuns entre os Portugueses no desenvolvimento de doenças respiratórias, tais como asma, doença pulmonar obstrutiva crónica (DPOC) e bronquiectasias.
2) Caracterizar o impacto de quatro mutações -711+1G>T, 711+3A>T, 711+3A>G e 711+5G>A localizadas na mesma sequência de consensus do intrão 5 e finalmente corrigi-las usando um único AON desenhado no nosso laboratório.

Para o primeiro objetivo, foi extraído DNA genómico de amostras de sangue dos indivíduos em estudo e, com técnicas de PCR (do inglês *polymerase chain reaction*) modificado, foram procuradas as nove mutações no gene da CFTR mais comuns para pacientes FQ portugueses.

Os resultados mostram que três das nove mutações mais comuns foram encontradas em pacientes com doenças respiratórias, onde a mutação mais comum foi a F508del, a mais comum na população mundial. Adicionalmente constatou-se que a G576A, descrita como sendo um polimorfismo, também parece estar envolvida na suscetibilidade de aparecimento de doenças respiratórias mesmo quando presente individualmente, sem que o outro alelo do indivíduo esteja mutado, porém este polimorfismo não foi encontrado entre a população FQ. Estes resultados sugerem que os polimorfismos no gene da CFTR podem estar envolvidas no aparecimento de asma, bronchiectasias e DPCO, e ainda que mutações num só alelo deste gene podem aumentar a suscetibilidade de um indivíduo portador desenvolver doenças respiratórias, sem que haja manifestação clínica típica de FQ.

Para o 2º objetivo, utilizámos um mini-gene de CFTR consistindo num construto de cDNA de wt-CFTR previamente clonado no plasmídeo pcDNA5, contendo dois intrões: o intrão quatro (IVS4) e o intrão cinco (IVS5) entre os exões 4 e 5 e os exões 5 e 6a, respetivamente, e dois marcadores (eGFP em N-terminal e Flag-tag no 4º *loop* extracelular) . Introduzimos cada uma das mutações de *splicing* em estudo por mutagénese direcionada neste mini-gene repórter. De seguida, cada um dos construtos foi clonado num vetor lentiviral para a produção de novas linhas celulares (experiência por concluir para comparação de resultados). Toda a caracterização apresentada ao longo deste trabalho, foi feita em células que expressavam cada um dos construtos por transfeção transitória. O impacto de cada mutação expressa pelas células foi caracterizado a nível do mRNA, expressão e localização da proteína na membrana. Resultados de PCR quantitativo (qPCR) e semi-quantitativo revelaram que as quatro mutações provocam a remoção total do exão cinco e que a mutações 711+1G>T e 711+5G>A diminuem drasticamente os níveis de transcritos normais, mais do que as mutações 711+3A> G e 711+3A>T. Com a técnica de Western blot (WB) verificou-se que a proteína resultante dos construtos com as mutações 711+1G>T, 711+3A>T, 711+3A>G e 711+5G>A é menor do que a proteína normal. Para além disso, quando caracterizadas por imunofluorescência, notou-se uma redução drástica dessas proteínas na membrana plasmática (PM), sugerindo um défice no processamento e possivelmente no tráfego da mesma. Na tentativa de corrigir todas as quatro mutações com o mesmo AON, este foi desenhado à distância de 20 bases a jusante do primeiro nucleótido do intrão 5, e verificou-se que um único AON aumentou o nível de transcritos normais produzidos e a expressão da proteína na PM.

Na sua globalidade, os resultados obtidos ao longo deste trabalho indicam que as mutações no gene da CFTR mais comuns encontradas entre portugueses poderão estar envolvidas no desenvolvimento de doenças respiratórias sem que haja manifestação clínica de FQ, e ainda que um único AON pode beneficiar pacientes com pelo menos uma das mutações estudadas.

Estudos futuros, expandindo a lista das mutações no gene da CFTR de pacientes com doenças respiratórias são importantes para potenciar e comprovar a associação entre a proteína CFTR e essas doenças. Além disso a determinação do genótipo e a caracterização do efeito das mutações ajudará a relacionar o genótipo com o fenótipo, melhorando assim o prognóstico tanto de pacientes com FQ como também a avaliação da suscetibilidade de um indivíduo portador desenvolver outras doenças respiratórias.

A correção das mutações de *splicing* em estudo por este AON precisa ainda de ser avaliada em sistemas celulares derivados de materiais de doentes *ex vivo* para tornar os resultados mais robustos e fisiologicamente relevantes. Com efeito são já extremamente encorajadores e comprovar em materias de pacientes a eficácia dum mesmo AON para corrigir 4 mutações de *splicing* diferentes, aumentará a sua potencialidade para uso futuro *in vivo* em pacientes com pelo menos uma destas mutações de *splicing*.

Palavras chave: Fibrose quística, mutações, splicing, CFTR, doenças respiratórias

List of abbreviations

3'ss	3' splice site
5'ss	5' splice site
AON	Antisense Oligonucleotides
ARMS	Amplification Refractory System
ATP	Adenosine Triphosphate
BI	Body-mass Index
BSA	Bovine Serum Albumin
CABVD	Bilateral Absence of Vas Deferens
CF	Cystic Fibrosis
CFBE	Cystic Fibrosis Bronchilial Epithelia
CFTR	Cystic Fibrosis Transmembrane Conductance Regulator
COPD	Chronic Obstruction Pulmonary Disease
C-terminal	Carboxyl Terminal
DB	Disseminated Bronchiectasis
DNA	Deoxy Ribonucleic Acid
dNTP	Dinucleotide Triphosphate
EMEM	Minimum Essential Medium Eagle
EMEM	Minimum Essential Medium Eagle
ENaC	Epithelium Sodium Chloride
ER	Endoplasmic Reticulum
ESE	Exonic Splicing Enhancer
ESS	Exonic Splicing Silencer
Ex	Exon
Fwd	Forward
gDNA	Genomic DNA
HBE	Human Bronchial Epithelia
HNE	Human Nasal Epithelia
IRT	Immunoreactive Trypsinogen

ISE	Intronic Splicing Enhancer
ISS	Intronic Splicing Silencer
IVS4	Intron 4
IVS5	Intron 5
MVCC	Mutation Varying Clinical Consequence
NBD	Nucleotide Binding Domain
NBS	New-born Screening
NMD	Nonsense-Mediated Decay
NPD	Nasal Potential Difference
N-terminal	Amino Terminal
PAGE	Polyacrylamide Gel Electrophorese
PCR	Polymerase Chain Reaction
PFA	Paraformaldehyde
PI	Pancreatic Insufficiency
PM	Plasma Membrane
PS	Pancreatic Suffiency
PTC	Premature Stop Codon
PVDF	Polyvinylidene
qPCR	Quantitative PCR
RFLP	Restriction Fragment Length Polymorphism
RNA	Ribonucleic Acid
RT	Room Temperature
RT-PCR	Semi-Quantitative PCR
RV	Reverse
SnRNA	Small nuclear RNA
SnRNP	Small Ribonucleoprotein
TMD	Transmembrane Domain
WHO	World Health Organizatio

Section I: Introduction

1.1. Cystic Fibrosis (CF) and CF Diagnosis

Cystic Fibrosis (CF) is the most common autosomal recessive disorder among Caucasians, affecting about 75,000 individuals worldwide, around 32,000 individuals in Europe and approximately 300 individuals in Portugal with an estimated incidence of 1:6000 individuals in Portugal.[1,2] In medieval folklore, infants with salty skin, were considered "bewitched" because they died at an early age. Nowadays it is known that these infants had CF.[3]

The frequency of this disorder is highly variable and is often a function of ethnic and geographic origin of the affected patients.[4]

CF is an inherited life-threatening disease, and it is caused by mutations in a single gene, the Cystic Fibrosis Transmembrane Conductance Regulator (*CFTR*), which encodes CFTR protein, a chloride and bicarbonate channel responsible for regulation of ion transport across the apical membrane at the surface of certain epithelia.[5–7] The hallmark of the mutational spectrum in the *CFTR* gene is the very high frequency of the F508del mutation in the Caucasian population, responsible for about two-thirds of all CF chromosomes.[8]

Clinically this disease is characterized by elevated levels of Cl^- in sweat, meconium ileus, pancreatic insufficiency (PI), low body-mass index (BMI), infertility in males mostly because of bilateral absence of *vas deferens* (CBAVD), and undescended testicles or hydrocele, amenorrhea in females, severe nutritional and pulmonary involvement.[3,6,7,9]Furthermore, patients have a chronic or a recurrent cough, producing mucoid and purulent sputum. Recurrent wheezing and pneumonia, digital clubbing, nasal polyposis and sinusitis may also occur as CF manifestations.[10,11]

In summary, CF is a multi-organ disorder. However, the dominant cause of morbidity and mortality is lung disease. There is a significant variability of CF symptoms among different patients, thus posing some difficulties in the clinical diagnosis.[12]

Diagnosis

The diagnosis of CF has traditionally relied on recognition of characteristic clinical symptoms and a family history of CF.[12] However, to confirm a diagnosis of CF evidence of CFTR dysfunction is required. Approaches to demonstrate the latter include: *(i)* two consecutive sweat tests showing a high Cl^-concentration in sweat (≥60mmol/L) *(ii)* two *CFTR* mutations previously described as CF-causing (mutations which are undefined or defined as a "mutation of varying clinical consequence" (MVCC), are not valid to confirm a diagnosis of CF); *(iii)* abnormal transepithelial Nasal Potential Difference (NPD) measurements; or *(iv)* absence or defective CFTR function determined in native colonic epithelium.[13–15] Although each test mentioned above is recommended, there is no need to perform them all. Thus, a diagnosis of CF can be hierarchically established (see Fig 1.1), being the sweat Cl^- the first one, secondly CFTR genetic analysis, and finally CFTR physiologic tests.[16] Therefore, when the concentration of Cl^- in sweat test is in the intermediate range (30-59mmol/L for infants less than six months of age, or 40-59 mmol/L for older individuals), *CFTR* genetic analysis is mandatory.[13,14]

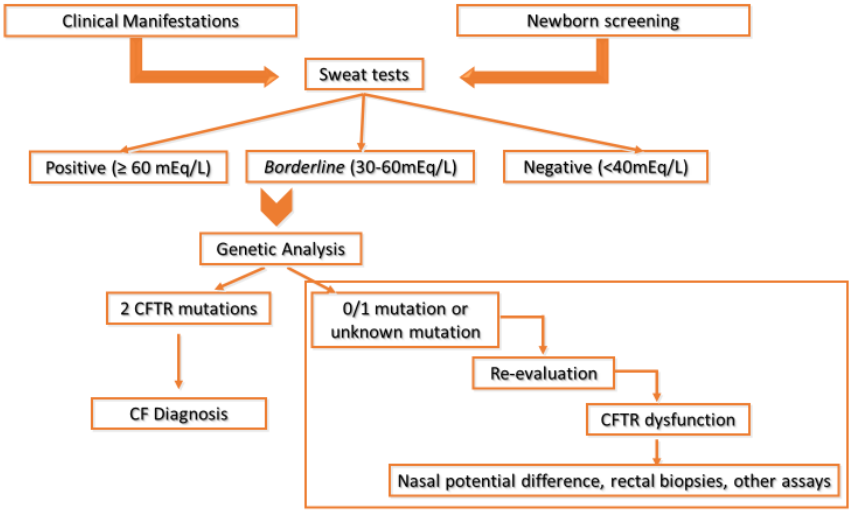


Figure 1.1 - Cystic Fibrosis Diagnosis Algorithm. The diagnosis of CF is firstly relied in clinical manifestations and in newborn screen results, secondly, it is necessary to have two positive sweat tests and finally two known CFTR mutations. If the results of sweat tests are inconclusive, for instance, borderline results, and if it is found only one CFTR mutation it is necessary to re-evaluate the results and perform functional assays. [Figure from MD Amaral, included with permission]

Several kits to detect *CFTR* mutations are commercially available (using a discrete group of mutations, with the 50 most frequent mutation of each region), and they are advantageous because are fast and relatively cheap, which will detect 99% of affected individuals in the most populations.[16–18] Though these commercially available kits are limited because the diagnosis will be missed if patients carry mutations which are not included in the kit.[16–20] To overcome this limitation, full CFTR gene sequencing is recommended, and it will detect most *CFTR* mutations, but, it may also detect novel mutations with unknown functional effect and undetermined CF disease prognosis. If such, it is necessary to confirm the CFTR (dys)function by electrophysiologic tests such as NPD and measurements of CFTR-mediated Cl^- currents in colonic epithelial tissue (rectal biopsies).[7,12,13,16] In addition, it is important to identify CFTR mutations in CF patients with none or just one mutation previously identified using other approaches such as for example DNA or mRNA analyse (routine technique used in our laboratory) and to characterize their effects *ex vivo* and *in vitro*, to establish better the diagnosis and prognosis of the CF disease.

Recently the prenatal population screening for maternal CF carrier status and the newborn screening (NBS) have been widely used in the USA and Europe, which is helpful because early diagnosis and treatment reduce symptoms, improve health, and lower costs associated with disease complications.[12] In Portugal NBS consists in a 1st test for immunoreactive trypsinogen (IRT), an assay for Pancreatitis-Associated Protein (PAP) in case of positive IRT, followed by a 2nd IRT and if all positive a final genetic analysis. All these strategies have resulted in increased life expectancy, as the predicted survival age of a CF patient in the past, was only six months,[21] and now the predicted age is 40 years old in the USA,[22]43.5 in the UK,[23]and 30.7 years old in Portugal according to CF adult follow-up in specialized centres.[24]

All these diagnosis methods have been advancing due to a better understanding of the *CFTR* gene, the protein structure and its function.[25]

1.2. CFTR: from Gene to Protein and Function

CFTR gene and expression

The *CFTR* gene was cloned in 1989 by using chromosome walking and jumping, and linkage disequilibrium studies.[26] The gene comprises 27 coding exons (showed in figure 1.2) spanning over 190Kb on the long arm of chromosome 7 (7q31.2), and the transcript is 6.5Kb.[27]

The expression of *CFTR* is very complex and involves multiple tissue-specific transcriptions start sites, alternative first exons and alternatively spliced transcripts. The development, the pathologic conditions, the cell type and tissue, regulate the expression of *CFTR*. The sites of the *CFTR* expression are the epithelial surface throughout the body, such as submucosal glands, airways (the site of developmental regulation of CFTR expression), pancreas, the crypt in the intestinal tract, sweat glands, vas deferens, salivary glands.[28,29]

The proper transcription of the CFTR gene gives rise a multi-domain glycoprotein of 1,480 amino acids with a molecular weight of approximately 170 kDa. This protein belongs to the superfamily of the ATP-binding cassette (ABC) transporters.[26,28,30]It is composed by five domains, as shown in Fig. 1.2: two transmembrane domains (TMD1 and TMD2, each comprising six transmembrane segments) which anchor the protein in the membrane and form the translocation pathway, two cytoplasmatic nucleotide binding domains (NBD1 and NBD2) which bind and hydrolyze ATP, and a central regulatory domain (RD).[28,31,32]

Function

The CFTR protein plays different roles according to its system/organ localization. For instance, in the intestine, pancreas and sweat gland secretory coil, CFTR plays a key role in fluids and electrolyte secretion, and in the sweat gland duct and the airway epithelia, it participates in fluid and electrolyte absorption.[33] It functions as a Cl^- and HCO_3^- channel and it also works as a regulator of other channels, for instance, Epithelium Sodium (Na^+) Channel (ENaC), that is downregulated by CFTR, potassium (K^+) channels and outwardly rectifying Cl^- channels (ORCCs),[33–35] or as a tumour suppressor gene in the intestinal tract.[36]

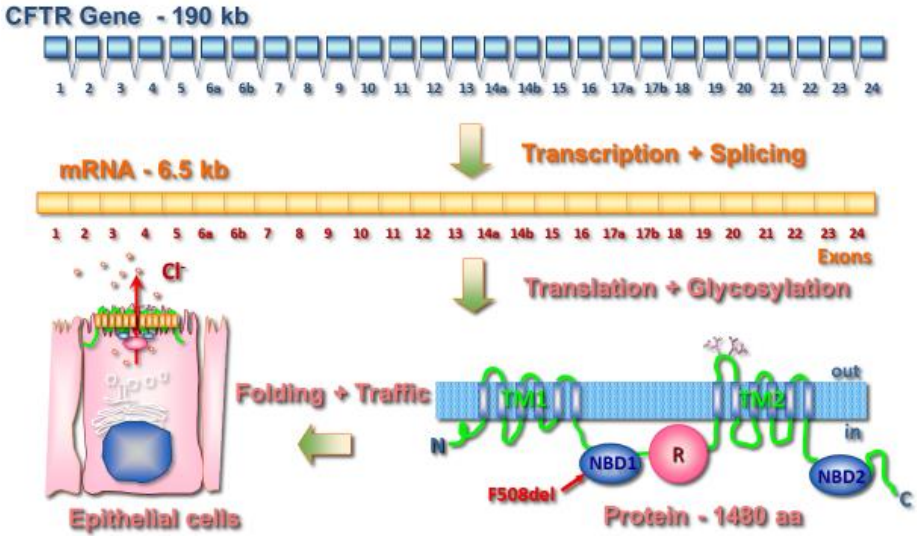


Figure 1.2: Schematic representation of transcription, translation and localization of CFTR. The CFTR gene is composed by 27 coding exons overspanning 190Kb. When every intron are skipped during the transcription (maturation of mRNA), it has 6.5Kb. The proper translation, trafficking and folding give rise a multi-domain glycoprotein composed by five domains: two transmembrane domain (TMD1 and TMD2), two nucleotides binding domain (NBD1 and NBD2) and one regulatory domain (RD). Both C- and N- terminus are localized in the cytoplasm. It is also composed by six extracellular loops and four intracellular loops. [Figure from MD Amaral, included with permission]

The production of defective CFTR protein results in abnormal transport of salt, not just due to abnormal Cl^- but also due to an enhancement of Na+ absorption, caused by ENaC upregulation thus causing a decrease in the water content and dehydration of epithelia of different organs, namely in the airways. This reduction causes thicker secretions in airway tract, which clog small airways, becoming a favourable environment for bacterial infections from the air (inhalation, for instance). Accumulation of the thick mucus leads to persistent infection and chronic inflammation. Because of chronic inflammation, the *bronchi* dilate, and their walls weaken, setting up bronchiectasis that results in further airflow obstruction. The cycle of airway obstruction, inflammation, and persistent infection leads to a progressive decline in lung function and eventually causes respiratory failure and lastly to death (Fig.1.3).[37–39]

When CF was first described, the incidence of death among children in infancy or early childhood was very high. However, progresses in symptomatic therapies, such as mucolytics to dissolve the thick mucus, antibiotics to treat or prevent infections, anti-inflammatory agents to ameliorate chronic inflammation, pancreatic enzymes to compensate exocrine pancreatic insufficiency, fat-soluble vitamins and high caloric intake to overcome the deficiency of vitamins and malabsorption of fats, respectively, have significantly improved the life expectancy of CF patients.[6,10,27,40]

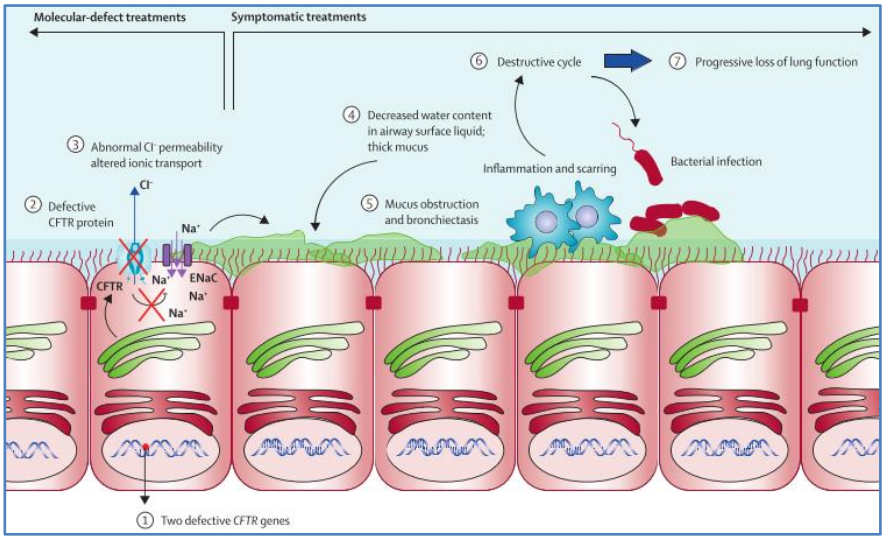


Figure 1.3: Pathophysiologic cascade in Cystic Fibrosis that leads to lung disease, the main cause of high mortality rate among CF patients. Normally, CFTR proteins are located on the surface of the epithelial membrane and act as Cl- channels that in turn regulate the ENaC (epithelial sodium channel). The complex interplay of these channels regulates the electrochemical gradient that allows appropriate airway surface liquid depth and mucus viscosity. When the CFTR protein is defective or missing (caused by mutations in both alleles) it alters the ionic transports (abnormal Cl^- permeability and high rate of Na^+ absorption into the cells), which leads to decrease of water content in the airway surface liquid, which in turn leads to production of viscous mucus. ENAC – Epithelium Sodium Channel Adapted from De Boeck & Amaral (2015)[6]

1.3. CFTR mutations: classification according to their functional defect and class-specific therapies

Until now 2,019 variations have been found,[41] with the following distributions: missense (39%); frameshift (16%); splicing (11%); nonsense (8%); large (3%) and in-frame (2%) deletions/ insertions; and promoter (1%); sequence variation (14%), unknown effect (6%).[27,40,41] However, only about 250 have been clearly defined as CF-causing, where approximately 20 mutations occur worldwide with frequency about 0.1% among CF patients, becoming very difficult to define the disease liability of such infrequent mutations.[5,27,40,41]

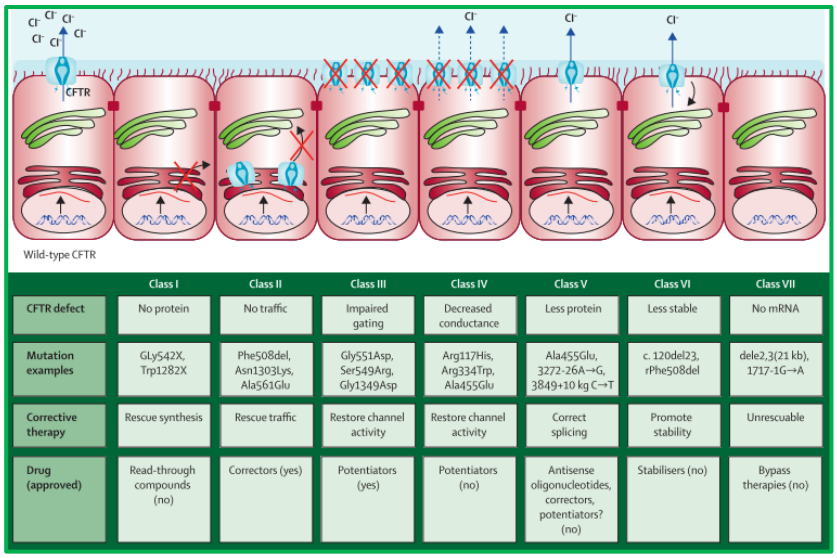
Although treatment advances over the past several decades have raised the median predicted survival age, the treatment costs of CF are very high, because of the amount of medicines taken daily by CF patients, which many drugs that prevent and treat pulmonary complications are costly.[7,10,42]

Consequently, a new approach that corrects the basic defect of CF, in this case at the molecular level is needed.

The basic defect in CF can be treated using small molecules to modulate defective CFTR protein and restore functional ion transport. There are already in the market compounds that restore CFTR trafficking and function such as, lumacaftor (VX809 – a corrector), ivacaftor (VX770 - potentiator), orkambi(VX809+VX770).[43,44] Ivacaftor increases the time that CFTR channel is open, allowing Cl^- ions to flow through the CFTR proteins on the surface of epithelial cells, and lumacaftor facilitate the trafficking of CFTR protein, thus allowing this protein to reach the membrane and transport Cl^-.[45–48]

However, these compounds do not restore all CFTR mutations currently described, being necessary a mutation-specific or mutation-class-specific approach to correct the protein defect. Indeed, CFTR mutations have been conventionally grouped into seven classes according to their functional defect (Figure 1.4):[5,12,27,40,48,49]

- Class I mutations – result in no protein production, mostly nonsense or premature stop codon (PTC) mutations, which results in a truncated mRNA, leading to its degradation by nonsense-mediated decay (NMD), an mRNA surveillance mechanism. In this class, CFTR production could be obtained by using read through compounds to mask the PTC, leading to the NMD scape and allowing the cell to produce a full-length protein[21,42].
- Class II mutations – affect CFTR protein traffic to the correct cellular localization as a result of protein misfolding and retention in the endoplasmic reticulum (ER), resulting in premature degradation of the protein.[50] The most common mutation, F508del belongs to this class. Small molecules, such as correctors can be used to correct this class of mutations, allowing the defective protein to reach the PM.[46,51–53] Recently, Orkambi (a mix of XV809 and VX770) was approved by the Food and Drug Administration for some of the patients with this class of mutations.[54]
- Class III mutations – lead to impaired gating of the CFTR channel, allowing the traffic of CFTR to the apical membrane, but causing poor regulation of the Cl^- channel. They result in a very poorly functional protein. Potentiators, such as VX770 are used to increase the CFTR channel gating, enhancing then the Cl^- transport.[27]
- Class IV mutations – cause a significant decrease in CFTR channel conductance, notwithstanding the CFTR protein is present at the apical surface. Potentiators could improve the CFTR function.[6,49,55]
- Class V mutations – lead to a major reduction of normal CFTR protein levels, mainly because of aberrant splicing. These mutations produce normal and aberrant mRNA and the proportion of mRNA variants may vary among patients in different organs of each patient. Patients with this class of mutations can beneficiate with potentiators that will enhance the function of the normal protein and RNA-based therapies that can enhance the normal mRNA production.[49,56–58] Recently VX770 has been approved for some of this class of mutations.[59]
- Class VI mutations – lead to a high turnover of CFTR at the cell surface, either by increasing CFTR endocytosis or by decreasing its recycling back to the cell surface. Stabilizers could be used for this class of mutations.[6,27]
- Class VII mutations – lead to a non-functional and rescuable protein because of large deletions or insertions. Mutations in this class cannot be pharmacologically rescued, and therefore the promising therapies are Bypass therapies, i.e., by activating other alternative Cl^- anion channel.[6,60]



	Class I	Class II	Class III	Class IV	Class V	Class VI	Class VII
CFTR defect	No protein	No traffic	Impaired gating	Decreased conductance	Less protein	Less stable	No mRNA
Mutation examples	GLy542X, Trp1282X	Phe508del, Asn1303Lys, Ala561Glu	Gly551Asp, Ser549Arg, Gly1349Asp	Arg117His, Arg334Trp, Ala455Glu	Ala455Glu, 3272-26A→G, 3849+10 kg C→T	c. 120del23, rPhe508del	dele2,3(21 kb), 1717-1G→A
Corrective therapy	Rescue synthesis	Rescue traffic	Restore channel activity	Restore channel activity	Correct splicing	Promote stability	Unrescuable
Drug (approved)	Read-through compounds (no)	Correctors (yes)	Potentiators (yes)	Potentiators (no)	Antisense oligonucleotides, correctors, potentiators? (no)	Stabilisers (no)	Bypass therapies (no)

Figure 1.4: Classification of CFTR mutations accordingly their functional defect, and its potential target therapies. Adapted from [6]

Patients with the same mutations have different responses to the same therapy, being crucial to study the response of these CFTR modulators in materials from each individual CF patients in a new strategy of personalised medicine. Thus primary cultures of human bronchial epithelial (HBE) cells and of nasal epithelial (HNE) cells as well as intestinal organoids produced from stem cells obtained from rectal biopsies directly from CF patients have been used as models for tests, such as measurements of the functional responses of CFTR and evaluation of drug efficacy.[61,62]

1.4. Class V mutations: splicing mutations and correction by RNA-based therapies

Alternative splicing is not always deleterious, is a fundamental element in eukaryotic gene expression that increases the coding capacity to the human genome. However, if an intronic or exonic variant disrupts canonical splice motifs or creates a new cryptic splice site, it will lead to aberrantly spliced mRNAs, usually encoding nonfunctional proteins. These, however, do not always replace totally the normal protein expression, as often some levels of normal mRNA splicing also occur in parallel with the abnormal one.[63–65]

The splicing process is very complex, and several reactions are taken in place. The spliceosome, the major effector of the splicing reaction, is a complex of hundreds of interacting proteins and small nuclear RNAs (snRNAs) including the five small nuclear ribonucleoproteins (snRNPs) U1, U2, U4, U5 and U6.[66]

First of all, U1 binds to the 5' splice site (5´-ss) by complementarity and U2 binds to the branch point. The triple snRNPs (U4, U5 and U6) complex moves-in to associate with the assembling spliceosome. Then, the U4 leaves the complex, allowing the replacement of U1 by U6 that interacts with U2 to bring the branch point into proximity to the 5´-ss. At this point, the first transesterification reaction cleaves the 5'-ss of the intron of the downstream exon and attaches it to the branch point. U5 then brings the 3'-ss of the upstream exon and 5'-ss of the downstream exon into proximity with each other, allowing a second transesterification reaction that cleaves the 3'-ss of the upstream exon (as shown in Appendix 1- Fig.1S). The splicing accuracy does not depend only on the mechanisms described above, but it also depends on more discrete elements, splicing regulatory elements which direct the splicing machinery to use the correct splice site. [66,67]

Patients carrying one splicing (class V) mutation can have mild CF disease, and most of the times there are considered pancreatic sufficient (PS). It is reported that 11% of the CFTR variations are known to be splicing defective. However, it is predicted that the proportion of these splicing defects can be

higher because of the false missense mutations that have not been identified as causing abnormal splicing.[40,68,69] In this way each nucleotide modification, including nonsense, missense and silent modifications, may potentially impact the splicing pattern, resulting in production of aberrant spliced transcripts.[70] Thus, it becomes crucial to study splicing mutations, which allow a better understanding of splicing processes and the mechanism of disease-causing, which may help to develop the most relevant therapeutic strategy and thus contributing to the increase of life expectancy of patients carrying these class of mutations.

When the alteration is in the invariant donor splice site, +1, it is most likely that the correct spliced transcripts will strongly be reduced.[53,71]

For clarity, when it is referred to intronic splicing mutations, for example, 711+1G>T or 3272-26A>G, the symbols "+" and "-" indicate nucleotides downstream and upstream of the exon/intron boundary, respectively.

RNA-based therapies

RNA-based therapies are therapies use RNA as a target to potentially tread diseases, caused by aberrant splicing. Targeting the RNA is an advantageous strategy because it avoids many of the risks and concerns associated with (DNA) gene therapy, such as random gene insertion. Furthermore, when an *in vivo* cellular basis is needed the presence within the cell of multiple different RNA processing pathways means that there is much scope for influencing its control at different levels. The dynamic nature of RNA turnover also implies that therapeutic interventions can be time-limited, dose titrated and modified according to response, adding further levels control.[66]

Novel therapeutic approaches to correct splicing mutations have been described using antisense oligonucleotides (AONs), that correct splicing defects, indicating that CF patients carrying such mutations may benefit from AON treatment.[57]

These AONs can be synthesised to be complementary and specific to a particular RNA sequence transcribed from the CFTR gene, meaning that only the RNA sequence of interest will be targeted. By designing AONs that bind to splice sites or to enhancer or silencer elements within the transcript, the splicing mechanism can be manipulated in a precise and reproducible way. Blocking splice sites and regulatory sequences prevents snRNPs and splicing factors such as SR proteins and hnRNPs from binding to the mutated site at the RNA transcript, allowing a directed exon skipping or inclusion, depending on the sequence blocked.[72]

Although RNA repair is transient and only modestly effective, studies of CFTR splicing polymorphisms suggest that 8% of normal CFTR message might be sufficient for normal lung function, whereas 5% is associated with relatively mild CF lung disease.[73,74]

1.5. Objectives of the present work

This project had two main goals, namely:

1) Firstly, to investigate the possible involvement of the CFTR in other respiratory diseases namely, COPD, DB and asthma.
2) Secondly, to characterize the impact of four mutations localized in the same splicing consensus- 711+1G>T, 711+3A>T, 711+3A>G and 711+5G>A and subsequently modulate them using an AON based strategy.

To achieve these goals we proposed the following specific tasks:

1.1. To screen non-CF individuals with other respiratory disorders for the most common CFTR mutations found in Portuguese CF patients. Control groups included: i) a group of patients known to have CF or suspicious of a CF diagnosis; ii) a group of carrier individuals; and iii) a group of healthy individuals with no respiratory phenotype.

2.1. To generate cells lines expressing CFTR minigenes carrying the splicing mutations 711+1G>T, 711+3A>G, 711+3A>T and 711+5G>A.

2.2. To characterize the impact of four splicing mutations (711+1G>T, 711+3A>G, 711+3A>T and 711+5G>A) at the mRNA, protein and functional levels.

2.3. To use an AON strategy for modulation of the deleterious effect of these mutations at the mRNA and protein levels;

Section II: Materials and Methods

Part I: Methods to analyse patients' materials

2.1. The screen of mutations in CF and non-CF patients

Before the studies, all individuals signed an informed consent. The patients' materials in this study were irreversibly anonymised as stipulated in artº 19, in the Portuguese law 12/2005 from 26th of January.

Initially, it was proposed to analyse the nine most common Portuguese *CFTR* mutations (panel obtained from World Health Organization (WHO), 2004 – shown in table 3.1) in individuals with suspicion of CF (who had two borderline sweat tests or just one identified mutation) and individuals with a CF familial history. However, for this study, the objective was extended to individuals who had two negative sweat tests and no CF familial history but have diagnosis of COPD, DB or asthma, to verify if the most common Portuguese CFTR mutations play a role in these respiratory diseases. It was also included healthy control, in order to discriminate better the CFTR sequence variation from CFTR mutations.

In order to achieve that, 249 Portuguese individuals divided into 4 four groups, namely, patients with suspicion of CF, individuals with CF sibling or with a CF familial history, non-CF with other respiratory diseases and healthy controls (individuals without a CF familial history, without clinical manifestations of any respiratory disease) were analysed.

Genomic DNA (gDNA) was extracted from blood samples of 249 Portuguese individuals, using the Wizard® Genomic DNA Purification kit following the manufacturers' instructions performed by Verónica Felicio and further analysed for 9 most common CFTR mutations (table 3.1) extracted from a report of word Healthy Organizaton.

The CFTR mutations were detected by Amplification Refractory Mutation System (ARMS), or Tetra-ARMS and Restriction fragment length polymorphism (RFLP) techniques, which were previously optimized in our laboratory.

The ARMS technique is a simple, rapid and reliable method for the detection of any mutation involving single base changes or small deletions. This PCR modification consists of two complementary reactions. The first reaction contains an ARMS primer specific for the normal gDNA sequence and cannot amplify mutant DNA and the second one contains a mutant-specific primer and cannot amplify normal gDNA. The genotype of an individual can then be determined by analysis of the amplification products. An individual with no mutations generates a PCR product only in the normal reaction, an heterozygous individual gives products in both reactions, and a homozygous mutant individual does so only in the mutant reaction.[75,76]

The tetra-ARMS technique is an ARMS modification using tetra-primers in the same reaction. Thus, the allele-specific amplification is achieved in a single PCR reaction using two outer primers and two allele-specific inner primers.[77,78] Then PCR reactions were run on a 1%-3% agarose gel to confirm amplification and product pattern. Mutation detection was confirmed by sequencing.

Part II: Methods to study splicing mutations in cellular systems

2.2. Generation of models to study splicing mutations in IVS5

2.2.1. Plasmid vectors

A *pcDNA5/FRT/eGFP/CFTR/IVS4art/IVS5/Flag* mini-gene was previously constructed in our lab (the insert will be called here as eGFP/IVS4_5/wt/flag or the respective mutants). The construct carried a complete wildtype (wt) CFTR cDNA, containing an artificial intron 4 (IVS4art) with 351 nucleotides of the 5' of the human IVS4 sequence joined with 351 nucleotides 3' of the human IVS4 sequence inserted between exon 4 and exon 5, and a full length of intron 5 (IVS5) sequence from human gDNA

inserted between exon 5 and exon 6a. It is also composed by eGFP in the N-terminal and flag in exon 15 (4th extracellular loop), which was previously inserted in our lab. This mini-gene was then used to generate four splicing mutations (711+1G>T; 711+3A>T; 711+3A>G and 711+5G>A) by-site directed mutagenesis.

Afterwards, each mini-gene was used to transiently transfect HEK293T cells and CFBE parental cells and was also used to stably transfect HEK FLP-In cells. The pcDNA5 mini-gene was also used to amplify the insert and clone into pLVX-puro to stably transfect CFBE parental cells.

2.2.2. Site-directed mutagenesis

Mutations on the same splicing consensus on IVS5 (711+1G>T, 711+3A>G, 711+3A>T and 711+5G>A) were introduced into *pcDNA5/FRT/eGFP/CFTR/IVS4art/IVS5/FLAG* using the KOD Hot Start DNA Polymerase Protocol (Novagen). Where 10ng of plasmid were mixed with 1.5mM of $MgSO_4$; 1X of 10X Buffer for KOD Hot Start DNA Polymerase; 0.2µM of each dNTPs; 0.6µM of complementary pairs of mutagenic primers (described in Appendix 1 - Table 2); 0.02U/µl of KOD Hot Start DNA Polymerase and nuclease-free water to perform 50µl of reaction, using the PCR program described in Appendix 1 - Table 3.

5µl of the reaction product were run by electrophorese in agarose 0.5% agarose gel to confirm the amplification of the plasmid. The PCR product was then incubated with *DpnI* (a restriction enzyme that hydrolyses methylated and hemimethylated DNA) for 1hour at 37ºC. After that, 200µl of competent bacteria were transformed with the hydrolysis product and grown on LB agar plates with 100µg/mL of Ampicillin (Sigma-Aldrich) followed by the extraction of the plasmid. The plasmids were then sent for automatic DNA Sequencing (Stabvida, Costa Caparica, Portugal) to confirm the introduction of each mutation. All these steps are in detail described in the section of appendices (Appendix 1: 1-3).

2.2.3. Cloning into pLVX-Puro

The inserts from each of the above pCDNA constructs were subcloned into pLVX-puro in order to generate stable cell lines. The constructs of eGFP/IVS4_5/wt; eGFP/IVS4_5/711+3A>G/flag; and eGFP/IVS4_5/711+5G>A/flag were cloned into pLVX-Puro using In-Fusion® HD Cloning Kit (Clonetch), and the constructs eGFP/IVS4_5/711+1G>T/flag; and eGFP/IVS4_5/711+3A>T/flag were cloned into pLVX-Puro using T4 ligase (ThermoFisher Scientific). Then lentiviral particles were produced by human cells and thus create stable cell lines.

The cDNAs of the constructs were amplified through PCR reactions from the *pcDNA5/FRT/eGFP/CFTR/IVS4art/IVS5/FLAG* with wt and the different mutants, creating 15bp extensions with XhoI local restriction sites in both C- and N- terminal (eGFP/IVS4_5/WT; 711+3A>G; and 711+5G>A), and creating 15bp extensions with XmaI local restriction in N- terminus and SpeI local restriction in C- terminus (eGFP/IVS4_5/711+1G>T; and 711+3A>T) with primers described in table 2.1.

Table 2.1: List f primers used to amplify cDNA to clone into pLV-Puro.

Name	Sequence 5'→3'	Amplified inserts
Plvx_XhoI_EGFP_Fwd	GGACTCAGATCTCGAATGGTGAGCAAGGGCGAGGA	WT, 711+3A>G and 711+5G>A
Plvx_XhoI_CFTR_Rvs	GAAGCTTGAGCTCGACTAAAGCCTTGTATCTTGCATCTC	
XmaI_GFP_CFTR_Fw	ATATATCCCGGGATGGTGAGCAAGGGCGAGGA	711+1G>T and 711+3A>T
SpeI_CFTRend_Rv	ATATATACTAGTCTAAAGCCTTGTATCTTGCATCTC	

Separately, 3µg of the cDNAs resulting from amplification with XmaI and SpeI primers were mixed with 3U SpeI (NZYTech) and 3U XmaI (ThermoFisher Scientific), 1X NZYbuffer U (10 mM Tris-HCl, pH 7.4, 100 mM KCl, 0.1 mM EDTA, 1 mM DTT, 500 µg/mL BSA, 50% (v/v) glycerol) and deionized water to make up the 20µl total reaction. This mixture was then incubated for 3h at 37ºC, followed by enzyme inactivation for 20min at 65ºC.

5µg of pLVX-Puro were linearized using 5U of XhoI (NEB) restriction enzyme mixed with 1X NEBbuffer 3.1 (100mM NaCl, 50mM Tris-HCl, 10mM $MgCl_2$, 100µg/ml BSA, pH 7.9) and deionized water to make up the 25µl total reaction. The reaction was then incubated for 3h at 37ºC, followed by enzyme inactivation for 20min at 80ºC, to create ends equal to the XhoI primers extensions, used to amplify the cDNAs of eGFP/IVS4_5/WT; 711+3A>G; and 711+5G>A, allowing recombination between the vector and the construct. 3 µg of pLVX-Puro was linearized using 3U of XmaI (ThermoFisher Scientific) and 3U of XbaI (NZYTech) mixed with 1X NZYbuffer U (10 mM Tris-HCl, pH 7.4, 100 mM KCl, 0.1 mM EDTA, 1 mM DTT, 500 µg/mL BSA, 50% (v/v) glycerol) and deionized water to make up the 20µl total reaction. . The reaction was then incubated 3h at 37ºC, followed by enzyme inactivation 20min at 65ºC, to create compatible ends of XmaI and SpeI primers extensions, used to amplify the cDNAs of eGFP/IVS4_5/711+1G>T; and 711+3A>T, allowing the ligation between the construct and vector. It is important to highlight that XbaI and SpeI generate compatible ends.

The linearized vector, as well as the cDNAs extended, were spin-column purified using the NzyGelPure kit (NZYTech). The In-Fusion cloning was performed mixing 2µl of the enzyme, 200ng of cDNA extended, 50ng of linearized vector and nuclease water free up to 20µl, and let it react 15mi at 50ºC. The ligation reaction was performed mixing 1U of T4 ligase (ThermoFisher Scientific), 213ng of cDNA extended and purified, 50ng of linearized vector and nuclease water free up to 20µl, and let it react 10min at 22ºC.

10 µl of the infusion product or the ligation product were then used for bacteria transformation as described in Appendix 1: 3.

'Colony PCR' was used to confirm the insertion of the construct into the pLVX. After confirming the insertion, the recombinant plasmid was extracted from the positive colonies using NZYMiniprep kit (NZYTech, MB010) as described previously in Appendix 1: 3, and the recombinant plasmid was then digested with specific restriction enzymes EcoRI (ThermoFisher scientific) to confirm the correct insertion of the gene into the vector. If it was correct, the recombinant plasmids were sent for sequencing in StabVida.

After confirming the correct insertion of the insert into pLVX-Puro, the new lentiviral vector was used to produce lentiviral particles to transduce CFBE cells.

2.2.4. Production of lentiviral particles

Lentiviral particles with pLVX/eGFP/IVS4_5/mut/flag or pLVX/eGFP/IVS4_5/WT/flag were produced in Human Embryonic Kidney 293T (HEK 293T) cells.

The cells were seeded at density $5x10^5$ per well on a 6well plate, in Minimum Essential Medium Eagle with L-Glutamine (EMEM, Biowhittaker®) supplemented with 10% of Fetal Bovine Serum (FBS, GIBCO Life Technologies) and incubated 24h at 37ºC and 5% CO_2. The day after the medium was changed 3h before the transfection.

The cells were transfected using Calcium phosphate transfection which involves mixing DNA with calcium Cl^- in a buffered saline/phosphate solution to generate calcium-phosphate-DNA co-precipitate, which is then dispersed onto the cultured cells. Calcium phosphate facilitates the binding of the condensed DNA in the coprecipitate to the cell surface, and the DNA enters the cell by endocytosis.

5µg of pLVX (with the desired construct – WT or Mut); 5µg of packaging plasmid pCMV-dR8.74.psPAX2 and 2.5µg of enveloping plasmid VsV-G/pMD2.G were mixed 1/10 of TE (1mM tris-Cl; 0.1mM EDTA; pH 7.60) and 2.5mM of $CaCl_2$ and this mixture was vortexed for 20 seconds and

incubated for 5 minutes. After that 250µl of HBS (50mM HEPES; 280mM NaCl; 1.5mM Na_2HPO_4) were added to another tube and while vortexing, the previous mixture was added dropwise and then incubated for 30 min. Afterwards, the mixture was added to the cells.

The cells transfected were incubated for 24h at 37ºC, 5% CO_2. The medium was then changed to EMEM supplemented with 10% FBS to remove the transfection reagent, and the cells were incubated for more 42 hours at 37ºC, 5%CO_2.

The lentiviral media were harvested, and the packaging cells were discarded. The lentiviral media were immediately used to transduce cystic fibrosis bronchial epithelial (CFBE 41o for short CFBE) cells or were stored at -80ºC for further use.

2.2.5. Generation of stably transduced cells: lentiviral infection

CFBE parental cells were plated in a 6 well plate at a concentration of $3x10^5$ cells per well in 2 ml of EMEM supplemented with 10% FBS a day before of infection with lentiviral media harvested from HEK 293T cells and were incubated at 37ºC, 5% CO_2.

The cells were infected with 1ml of lentiviral media and 1ml of EMEM supplemented with 10% FBS with 8µg/ml of Polybrene (Hexadimethrine bromide, Sigma-Aldrich) infection enhancer. The plate was centrifuged at 220rpm for 1h at 25ºC and then incubated for 24h at 37ºC, 5% CO_2. The medium was then changed 24 hours after the transfection to EMEM supplemented with 10% FBS. 42 hours post infection the media was changed to EMEM supplemented with 10% FBS and 2.5µg/ml of Puromycin, which selected the infected cells and killed the non-infected cells. Afterwards, the media was replaced 24 hours after the first selection with increased concentration of Puromycin (5µg/ml). The medium was then changed every 48 hours, and the cells were kept in culture at 37ºC, 5%CO_2.

To analyse the expression of the desired proteins in the stably transduced cells we performed biochemical analysis as described in section 2.4.

The cells were then sorted by flow cytometry to achieve a homogeneous expression of the protein.

2.3. Cell culture

The characterization of the splicing mutations was made in different cell types and different conditions.

2.3.1. Cell lines and culture conditions

Human Embryonic Kidney 293T (HEK 293T) parental cells were used for transient transfections and to produce wt and the mutants (711+1G>T, 711+3A>T, 711´3A>G and 711+5G>A) lentiviral particles. These cells were cultured in EMEM supplemented with 10% of FBS.

Human Embryonic Kidney293 Flp-In (HEK293 Flp-In) parental cells (Invitrogen, Carlsbad, CA) were used for stable transfections and were cultured in EMEM supplemented with 10%, selection antibiotic Zeocin 100µg/ml (Sigma-Aldrich, St. Louis, MO) before stable transfections and Hygromycin after transfections. The Flp-In system allows insertion of the construct in only a single integrated FLP recombination target (FRT) site.

CFBE parental cells, were cultured in EMEM supplemented with 10% FBS. These immortalised cells were developed from bronchial epithelial cells from an F508del-CFTR homozygous CF patient and did not express endogenous CFTR because it loses the expression with passages.

All cell lines were maintained at 37ºC in a humidified atmosphere of 5% (v/v) CO_2. All cells were tested for mycoplasma infection, being mycoplasma free.

2.3.2. Transient transfections

HEK293T cells were submitted to liposomal transfection. The liposomal transfection, commonly known as lipofection, is based on the ability of cationic lipidic to form unilamellar liposomes, which adsorb nucleic acids molecules to their surface and are capable of being internalised by the cells. We used Lipofectamine 2000 (Invitrogen) to transiently transfect HEK 293T cells with each pcDNA5 minigene

For that $5x10^5$ cells were grown in a 12-well plate 24 hours before the transfection, and 60%-80% of confluence was desired for transfection. Lipofectamine 2000 (3µl) and 1µg of DNA (pCDNA5 plasmids with the different constructs) were separately incubated for 5min in 50µl of OPTIMEM (Invitrogen), then mixed and allowed to incubate for 15min at RT. The mixture was then added to the cells in EMEM without FBS.

The cells were incubated for 48 hours at 37ºC in 5% CO_2, and the RNA and protein were extracted to perform RT-PCR, qPCR and Western blot, respectively as described Appendix 1: 1, 2 and 4 and in this section in 2.5.1, respectively.

2.3.3. Stable transfections in HEK Flp In cells

Lipofectamine 2000 was used to stably transfect HEK Flp In cell. A vector containing the flipase that allow the directed recombination, pOG44, was cotransfected with each desired constructs (pCDNA5/FRT/eGFP/IVS4_5/WT or mut/flag) in proportion 3:1 for a total 2µg of DNA, cell selection started 48h after transfection by changing the medium to a medium supplemented with 100µg Hygromycin B.

2.3.4. Treatment with an AON

The AON was synthesized by Integrated DNA Technologies, Inc. (IDT, Leuven, Belgium). The AON was designed 20 base pairs of distance from the first nucleotide of IVS5 with the following sequence: mA* mT* mC* mT* mT* mT* mT* mA* mG* mG* mC* mA* mC* mT* mA* mT* mT* mG* mT* mT*. The letter 'm' represents an O-methyl modification at the second position of a sugar residue, and the asterisk represents a phosphorothioate modification of the backbone. This AON was used to transiently transfect cells by Lipofectamine 2000 (Invitrogen).

48h after transfections mRNA was extracted to perform semi-quantitative and quantitative analysis by RT-PCR and qRT-PCR, respectively as described in the appendix section. Biochemical analysis were also performed by Western blot and Immunofluorescence to characterize the impact of splicing mutations under study.

2.4. Biochemical

2.4.1. Western Blot

Cells were washed twice with cold PBS and lysed with sample buffer (1.5% (w/v) SDS; 0.01% (w/v) bromophenol blue; 5% (v/v) glycerol; 0.05 dithiothreitol (DTT); 0.095 M Tris, pH 6.8). DNA was sheared using benzonase 25U/mL (Sigma-Aldrich) in the presence of 2.5 mM of $MgCl_2$.

Total protein was quantified using the Bradford assay. Briefly, 10µl of the protein extract was added to 990µl of BioRad Protein Assay Reagent (Bio-Rad, 500-0006EDU) diluted in water (200µl:800µl), incubated for 5min. The absorbance was measured at 595nm using the spectrophotometer (Jasco V-560 UV/Vis spectrophotometer). A regression equation for protein concentration was determined using bovine serum albumin (BSA) as a standard protein.

25µg of protein were loaded on each lane for SDS-polyacrylamide gel electrophoresis (PAGE) on 7% separating and 4% stacking gels, and run at 100 V for three hours in a 1X running buffer (25mM Tris; 192mM Glycine, 0.1%(w/v) SDS; pH 8.3, Bio-Rad). Subsequently, proteins were transferred onto

Polyvinylidene (PVDF) membranes (Millipore) at 400mA for 1.5h in a 1X transfer buffer (25mM Tris, 192 mM Glycine; pH 8.3, Bio-Rad).

After the transfer, the membrane was blocked 5% (w/v), non-fat milk in phosphate buffered saline (PBS, NaCl 137 mM; KCl 2.7mM; KH_2PO_4 1.5mM; Na_2HPO_4 6.5mM, pH 7.4) containing 0.1%(v/v) Tween (PBS-T) for 1h. The membranes were then probed overnight at 4ºC with primary antibodies (596 anti-CFTR mice (1:3000 (BD Transduction Laboratories®) and anti-calnexin mouse (BD Transduction Laboratories®) - as a loading control) diluted in 5% (w/v) skimmed milk in PBS-T. The membrane was washed three times during 10 min in PBS-T, followed by incubation for one hour at Room Temperature(RT) with horseradish peroxidase-conjugated secondary anti-mouse IgG antibody (BIO-RAD, goat, 1:5000) in 5% (w/v) skimmed milk-PBS-T and washed another three times during 10 min with PBS-T.

Chemiluminescent detection was performed using Chemidoc XRS+ analyser (BIO-RAD), and the signal was detected with the Clarity Western ECL substrate (BIO-RAD). Finally, the quantification was performed with the ImageLab software (BIO-RAD).

2.4.2. Immunofluorescence

Cells were grown on coverslips in 24-well plates or 8well- Lab-Tek (Sigma-Aldrich) coated with 0.001% (w/v) poly-L-lysine (Sigma) to reach 40-60% confluence was reached.

48hours after transient transfections the cells were washed once with PBS++ (PBS with components described in 2.4.1 enriched with 0.7mM of CaCl and 1.1 mM of $MgCl_2$) and incubated with mouse anti-flag (Sigma-Aldrich) diluted in PBS++ (1:500) supplemented with 1% (w/v) BSA for 1h at 4ºC. Afterwards, the cells were rinsed three times with PBS++ and fixed with 4% (w/v) Paraformaldehyde (PFA) for 20 min at 4ºC. The cells were then washed three times and incubated with a rabbit anti-mouse Cy5 secondary antibody (Life Technologies) in PBS++ (1:500) supplemented with 1% (w/v) of BSA, one hour at RT. After, the cells were washed three times with PBS and incubated one hour in the dark with Hoechst 33342 Fluorescent Stain Life (Life Technologies). The coverslips were then mounted on glass slides with Vectashield mounting medium (Vector Laboratories) and sealed.

Immunofluorescence staining was observed in the Leica DMI 6000B fluorescence microscope, which was also used to acquire the images.

2.5. **Statistical analysis**

Results are expressed as means ± S.E.M for *n* observations. Students t-test (GraphPad Prism software) for paired and unpaired samples was used as appropriate. Differences were considered statistically significant when $P<0.05$.

Section III: Results and Discussion

3.1. Analysis of Patients

3.1.1. Patients' screen studies

This study was carried out in order to determine whether patients with non-CF respiratory diseases have at least one mutation in the CFTR gene. Previous studies have shown that patients with, CB/AVD,[79] asthma[80], disseminated bronchiectasis (DB) and idiopathic pancreatitis with unknown etiopathology[81,82,83] have at least one CFTR mutation. However, there are also some studies showing that the incidence of CFTR mutations in patients with these non-CF respiratory diseases,[84] is not different from that in the general population,[85] thus being contradictory. This study will help to clarify these contradictory findings.

Herein, we tested the occurrence of the 9 most common Portuguese CFTR mutations (outlined in Table 3.1) in non-CF respiratory diseases Portuguese patients.

Table 3.1: List of the 9 most frequent mutations in Portugal adapted from[86]

Mutation	Incidence
F508del	59.31%
A561E	3.24%
R334W	2.83%
R1066C	2.83%
G542X	2.63%
N1303K	1.82%
G576A	1.42%
711+1G-T	1.01%
3272-26 A-G	1.01%

For that purpose, gDNA samples were obtained from 249 individuals including (see Table 3.2): 1) patients with non-CF respiratory diseases (136); 2) patients with CF or with suspicion of a CF diagnosis (30); 3) CF carriers (32); and 4) a control group of 51 healthy individuals with no respiratory phenotype. These gDNA samples were screened for the above mutations. Samples from the above 2-4 groups were included in this study as they were used to validate the pattern of the bands in the ARMS and Tetra ARMS PCRs used to determine each patient genotype (see Methods, section 2.1). In addition, they will be included in further studies aimed to correlate the occurrence of CFTR mutations with the severity of respiratory diseases and with the CFTR-mediated ionic transport. Whenever there was no clear genotype identified, samples were sent for sequencing to confirm the mutation(s).

The results revealed that 66 (26.5%) individuals had at least one CFTR allele mutated We were able to find only 7 out of the 9 most common Portuguese CFTR mutations described in Table 3.3, namely: F508del, 3272-26A>G, G542X, R334W, G576A, A561E and R1066C. However, we were also able to identify other 7 mutations that were not included in our cohort, namely, I148N, G85E, P205S, Q1100P, F1052V, V332Q and 3171del C. From the 14 different mutations that we found, F508del was the most common, and it was present in 62% of individuals with at least one mutation, and 14.6% of these individuals were homozygous for this mutation.

Looking into non-CF sub-group with respiratory diseases, the DB patients presented the highest incidence of the F508del mutation (6.5%), followed by the COPD patients (4%).

3.1.2. Genotype-clinical status Correlation

It is important to refer that asthma, COPD and DB are also present in some CF patients who are characterized by pancreatic exocrine dysfunction, increased sweat electrolyte[81], however, herein the patients with asthma, COPD and DB did not have any clinical manifestation which would be suggestive of a CF diagnosis.

Among the 249 individuals, 136 had non-CF respiratory diseases namely: DB- 31; COPD- 51; and asthma- 54. In this group, we identified the mutations G542X and F508del and a polymorphism G576A in 7 individuals (1 mutation in each individual).

We observed a variation regarding the incidence of mutated alleles for different pathologies within this group of individuals with non-CF respiratory diseases. Accordingly, the highest incidence was observed in COPD patients, with 4.9% of mutated alleles, followed by DB with 3.2% of mutated alleles (Table 3.2). The lowest incidence of CFTR mutations was observed among the patients with asthma, with only 0.9% of mutated alleles. Not surprisingly the frequency of CFTR alleles with mutations is higher in CF patients group (91.7%) than in the other groups.

Table 3.2: Summary of mutations identified in different respiratory illness (non-CF) and CF patients.

	Individuals	Total of individuals with at least one mutations	Total of alleles with mutations	% of alleles with mutations
Asthma	54	1	1	0.9%
DB	31	2	2	3.2%
COPD	51	4	5	4.9%
CF	30	28	55	91.7%
CF carrier	32	31	31	48.4%
Healthy control	51	0	0	0.0%

A more detailed analysis of the mutations identified and the patient's genotypes showed that The COPD group had the highest incidence of mutated alleles, in which 5 mutated alleles were found in 4 individuals and the F508del mutation was again the most common mutation (Table 3.2). A particular case was a 52-year old female with COPD who presented one severe mutation in one allele, the G542X mutation and one polymorphism, G576A in the other allele. This patient had two negative sweat tests and 79% FEV1 (Forced Expiratory volume in one second) and no digestive disease. The G576A polymorphism was not found in CF patients, only in CF carriers and COPD patients. In the other two respiratory diseases (asthma and DB), only F508del was found. Despite the G576A being described as a sequence variation, it can have some effect giving rise to another clinical status such as COPD, thus indicating that polymorphisms can also be involved in the etiopathogenesis of at least some cases of COPD. Stating that for future studies it is important to include polymorphisms in the cohort of mutations. In fact, in one study of incidence of mutations in Italian patients with DB has identified three rare DNA polymorphisms, and one of these three was the G576A polymorphism.[81]

All CFTR mutations identified here are CF-causing disease except G576A, described as a polymorphism in the CFTR database,[87] but its influence is under study because this variation is associated with several respiratory disorders.[88] In CF it has been shown that the G576A polymorphism can cause disease when in complex allele with the R668C mutation.[89,90]

In this study, more seven mutations were found when some samples were sent for sequencing, but interestingly these mutations were not found in patients non-CF respiratory disease (Table 3.3). These mutations were presented in about 0.05% of the overall population in the analysis. The I148N

mutation, where asparagine replaces isoleucine in position 148 of the exon 4 is not frequent among CF patients, and its severity is unknown,[87] but it has been described in one CF Portuguese patient in complex allele with 6 repetitions of 5T in intron 8 (IVS8-6(5T)).[91] The other mutations have been described in the Hispanic population,[92]Brazilian (Porto Alegre),[93] Spanish,[94] and Palestine residents.[95]

In this study, some mutations did not show the same incidence as in the reference table. Some examples are the A561E and the 3272-26 A>G mutations. The A561E mutation, in which glutamic acid replaces alanine at position 561 of the CFTR polypeptide had the smallest incidence, being present in 3% of the individuals with at least one mutation and 0.8% in overall individuals. This mutation was first described in a male nine years old Portuguese patient with moderate lung disease, but pancreatic insufficiency and is described as a second widespread mutation in Portugal present in 3% of the overall population, where so far four patients were homozygous for this mutation.[41,96] The 3272-26 A>G was the second more frequent mutation in this study present in 8 patients where one was homozygous for the mutation. This mutation was first described in a compound heterozygous with the severe mutation W846X, but the patient presented a very mild phenotype, with moderate pulmonary disease and was pancreatic sufficient. This mutation occurs in intron 17a creating an alternative splice acceptor site that produces an aberrant transcript that includes 25 nucleotides from intron 17a, causing a frameshift leading to a premature stop codon.[41,97]

Looking only at the results from CF patients analysed in this study, they are in agreement with results found in a retrospective study of CF patients in Portugal,[91] nevertheless, our study includes other groups and not exclusively CF patients, so it can explain the differences in incidence and mutations frequency that we found.

Table 3.3: Association between the number of mutations identified and Genotypes of CFTR found among each group analysed, CF patients (CF), patients with other respiratory diseases (Asthma, DB and COPD), CF carriers and healthy controls individuals.

	Asthma	***DB***	***COPD***	***CF***	***CF carrier***	***Healthy control***
Number of identified mutations	1	1	3	13	9	0
Individuals with 2 identified mutations	0	0	1	27	0	0
List of Genotypes	F508del/?	F508del/?	G542X/G576A G576A/? F508del/?	F508del/F508del F508del/3272-26 A>G F508del/R334W F508del/G85E F508del/P205S F508del/A561E F508del/G542X F508del/R1066C F508del/F1052V F508del/V332D 3272-26 A>G/3272-26 A>G 3272-26 A>G/3171 delC G542X>I148N G85E/G85E Q1100P/R334W R334W/?	F508del 3272-26 A>G A561E G542X G576A G85E I148N R334W V332D	----

In this study, 7 out of 136 individuals with non-CF respiratory diseases had mutations in at least one allele. Moreover, in 51 healthy control individuals, nobody had mutations in the *CFTR* gene, and the incidence of carries among Portuguese is 1 in 25 individual, so, it can be accepted that some of the most common Portuguese mutations can have some liability in other respiratory diseases. In our study, COPD has the highest CFTR mutations incidence, and asthma had the lowest. However, studies by other authors where the incidence of CFTR gene mutations and unclassified variants in COPD, DB and asthma in children and adults were investigated, it was found that asthma had the highest incidence of CFTR mutations and COPD had the lowest incidence,[98,99]. In a Greek study including patients with COPD, DB and asthma it was found some CFTR mutations [98] and any of these mutations were found in our study. This disagreement regarding the mutations could be because in our study we only searched for the most common Portuguese mutations and in the Greek population study, the screen was performed in all 27 exons of the CFTR gene. Other studies revealed that congenital mutations in the CFTR gene do not enhance the probability of having CFTR dysfunction in patients with COPD[84,100] and chronic bronchitis[85], however, in this study, we have shown that CFTR mutations can have an involvement in causing these non-CF respiratory diseases. But functional studies need to be done to confirm our results.

3.2. Study of splicing mutations in cellular systems

3.2.1. Generation of the models to study splicing mutations in IVS5

The second part of this project was to characterize the impact of four mutations located in the same splicing consensus, the 711+3A>T, which was first described in a Spanish family[101] and we also identified it in a Portuguese patient,[102] the 711+1G>T, first found in a Canadian family in compound heterozygosity with the 621+1G>T splicing mutation,[103] the 711+3A>G, first described in two families of Albanian origin[104] and the 711+5G>A which has also been previously identified in a Italian family.[105]

During the screening of mutations in samples of patients with the nine reactions technique developed in our laboratory[102] and gDNA sequencing, one novel splicing mutation was identified, the 711+3A>T.

Herein, we aimed to characterize the impact of the above mutations (each position showed in Fig.3.1) at the mRNA and protein levels. For that, we generated cellular models that enabled us to study the splicing defects caused by these mutations.

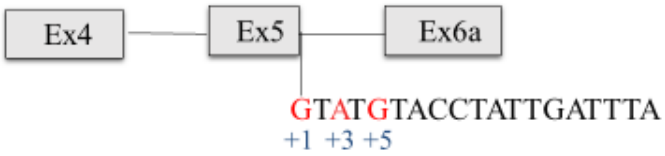


Figure 3.1: Schematic representation of the splicing mutations located in same splicing consensus. These mutations are located in the 5´´ splice site of intron 5.

A minigene approach was chosen to explore the impact of all these four mutations. So, using a minigene containing wt-CFTR cDNA with artificial intron 4 between exon 4 and exon 5 and intron 5 between exon 5 and exon 6a that already existed in the lab, *pcDNA5/FRT/eGFP/CFTR/IVS4art/IVS5/Flag*, we generated each splicing mutant by site-directed mutagenesis (see Methods).

These five minigenes were then used to transiently transfect both HEK 293T cells and CFBE cells. It was hard to generate stable HEK FLP In cells with these minigenes, due to an interference in the *Hygromycin B* coding sequence which was not allowing the survivor of cells when hygromycin was added in the media, as outlined in the methods (2.4.3). We then cloned the insert from pcDNA5 into PLVX-puro, a lentiviral vector to generate CFBE cells stably expressing each four constructs for further comparisons and characterization (this is under process and is not completely done). So, the results shown in this work were obtained in Cells transiently expressing pcDNA minigene.

3.2.2. Impact of the splicing mutations at the mRNA and protein levels

RT-PCR and qPCR: mRNA level

The impact of the splicing mutations was studied at the mRNA level, and for that purpose, HEK 293T cells and CFBE cells were transiently transfected with each minigene,

Firstly, the transcripts obtained from HEK293T cells transiently expressing each minigene were semi-quantified by RT-PCR. Primers that amplify the CFTR cDNA from exon 3 to exon 6a were used, and it was observed a strong band in wt-CFTR minigene, corresponding to the normal inclusion of exon 5 and no detectable alternative band (Fig.3.2). However, in all four mutations an alternative lower band was detected, corresponding to the transcripts with total skipping of exon 5 (Fig. 3.2) and a wt transcript was also detected. Although this analysis was semi-quantitative, it was possible to observe that the proportional intensity of wt transcripts was different among the four splicing mutations, being more

intense for mutants with the longer the distance from the canonical site. Accordingly, the 711+1G>T construct had no detectable wt transcripts.

It was shown that the skipping of exon 5 does not allow the occurrence of a premature stop codon because the joining of the exon 4 and exon 6a is in-frame.[106]

These data validated the experimental system and recapitulated the CFTR IVS4/5 splicing in patients, thus providing a good model for further studies.

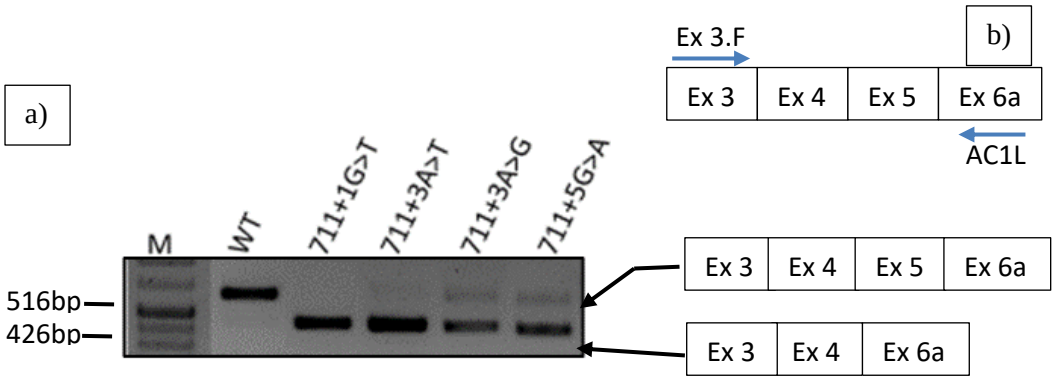


Figure 3.2: Characterization of CFTR transcripts of HEK 293T cells expressing the wt and all four mutants minigene by RT-PCR. The transcripts from HEK293T cells expressing wt or 711+1G>T minigene presented only one detectable band with different molecular weight (a), for each minigene, being more weighted the wt transcripts with the inclusion of exon 5 and the mutants with the lowest size corresponding to the exon 5 exclusion, as shown in the lower panel (b). The set of primers used for this characterization are shown in upper panel (b). Cells expressing the 711+3A>T, 711+3A>G and 711+5G>A minigene, presented two bands: one corresponding to wt transcripts, and the other one corresponding to aberrant transcripts with no exon 5.

The transcripts were then quantified by quantitative RT-PCR (qRT-PCR), and it revealed differences in abundance of WT transcripts among all mutations as previously observed in the semi-quantitative analysis (Fig. 3.3). However, in the quantification analysis, samples from cells expressing the wt minigene also had detectable levels of transcripts with no exon 5 (~5%) and those from cells expressing the 711+1G>T mutation minigene had a few wt transcripts (3%) – as shown in Fig. 3.3, that was not detectable by semi-quantitative RT-PCR. Cells expressing the 711+3A>T minigene seem to have the highest amount of wt transcripts when compared to the cells expressing the other three mutations in the study. Interestingly, we detected a diminished amount of wt transcripts caused by the 711+5G>A mutation, 7.4%±5. In our lab, another mutation in the +5 position (2789+5G>A) in intron 14b has been previously studied, and it was observed that mutants expressing the 2789+5G>A minigene have 56% of wt transcripts.[57]

The difference between the two techniques is that the qPCR technique is more sensitive and can detect lower expression levels of transcripts compared to the RT-PCR technique. These data illustrate the importance of performing both techniques for transcripts characterization.

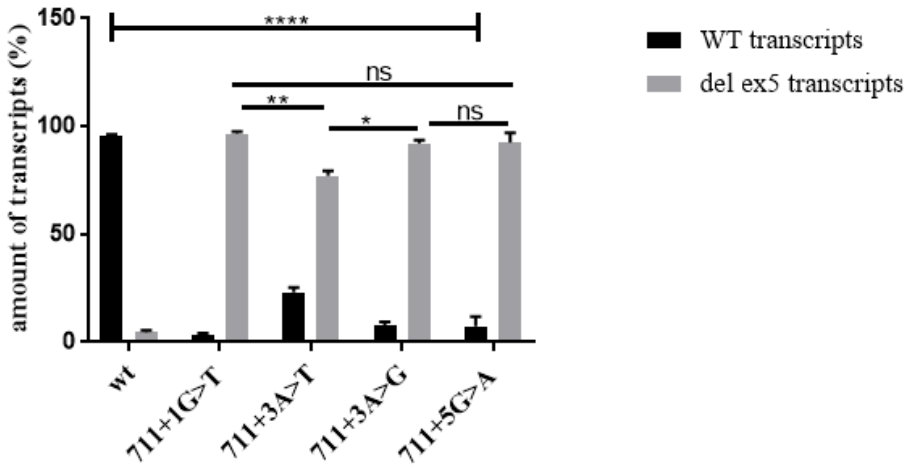


Figure 3.3: Quantitative analysis of CFTR transcripts with exon 5 inclusion (WT transcripts) and exon 5 exclusion (del ex5 transcripts) from HEK 293T cells transiently expressing different minigenes (WT and mutants). 2way ANOVA Multiple comparisons were used to compare wt and del ex5 transcripts among the various splicing mutations and WT minigene. The statistical differences are expressed with multiple *P values*, being P=0.015 (*), P=0.0015 (**), P<0.0001 (****) and P≈0.9 (ns). Data are expressed as mean ± SEM of three independent experiments.

Results obtained in transient transfections of HEK 293T cells were then confirmed using mRNA extracted directly from native tissues of individuals with this mutation. For that purpose, cDNA was synthetized from mRNA extracted from small pieces of lung of individuals with no mutation (control) and from organoids (a 3D-structure grown from stem cells extracted from rectal biopsies of individuals) from two patients with the 711+1G>T/F508del genotype.

Interestingly, in individuals with the 711+1G>T/F508del genotype, ~50% of the transcripts are aberrant and the other ~50% are normally spliced transcripts, suggesting that the normally spliced transcripts result from the F508del allele and the transcripts with no exon 5 result from the 711+1G>T allele (Fig. 3.4). These data show that in these individuals the mutation 711+1G>T originates only aberrant transcripts. In parallel, no aberrant transcripts were detected in the healthy control.

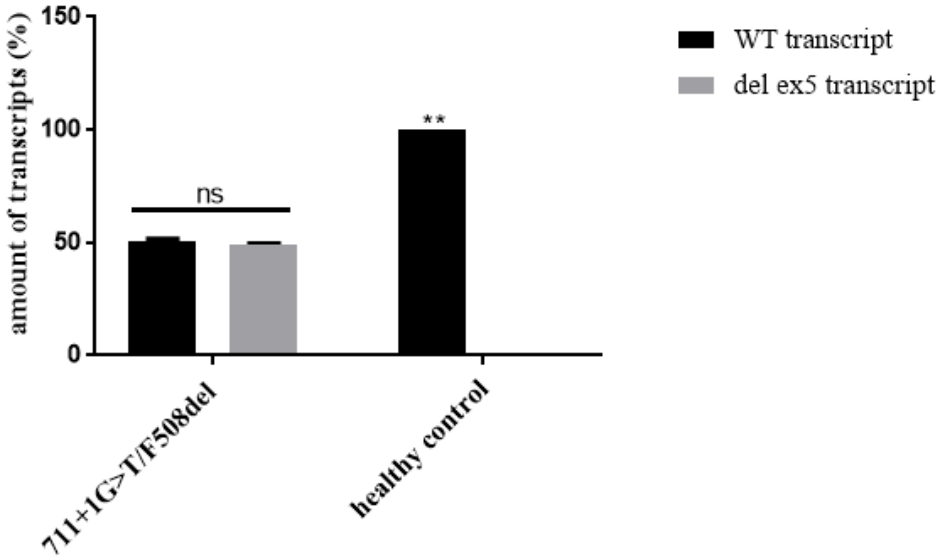


Figure 3.4: Quantitative RT-PCR analysis of CFTR transcripts (normal and aberrant transcript with no exon 5) from individuals with cystic fibrosis (711+1G>T/F508del) and healthy control. Multiple *t*-tests were used to see the differences between the amount of wt and transcripts with no exon 5 (del ex5 transcript). The amount of wt transcripts and aberrant transcripts (del ex5 transcript) are not statistically different in CF patient (ns) with $P=0.028$, and they are statistically different (**) in healthy control, with $P<0.0001$. Data are expressed as mean ± SEM of four independent experiments for the 711+1G>T/F508del and two independent experiments for the healthy control.

In summary, the results obtained by qRT-PCR and RT-PCR suggested that the 711+1G>T is a severe splicing mutation, while the 711+3A>T and 711+3A>G splicing mutations could be considered milder mutations because of the amount of wt transcripts produced in cells expressing the splicing mutations, since it has been shown that patients with 5% of normal CFTR transcripts have a significantly milder disease.[73] The 711+5G>A mutation is described in the CFTR database, as causing PI when combined with other mutation that also causes pancreatic insufficiency,[87]showing the severity of this mutation.

The quantification of transcripts obtained in CFBE cells transiently expressing the WT and all splicing minigenes, also showed that all mutations strongly reduced the levels of correctly spliced transcripts (data not shown). Additionally, it revealed that the 711+1G>T did not originate any correctly spliced transcripts, giving slightly different results from that obtained in HEK 293T cells. These data suggest that different cell types can process splicing differently, being thus important to confirm all results obtained by the minigene approach by direct mRNA analysis in patient's samples.[71]

All these mutations alter the WT donor site,[107]thus leading to aberrant splicing. However, the mechanisms of the splicing disruption in all these mutations are not entirely understood.

Using the Human Splicing Finder (HSF) software to explore the elements that regulate the splicing, it was possible to identify three splicing elements just downstream of the invariant G (GT- at the 5' end of the intron or the donor splice site), namely, 2 intronic splicing enhancer elements (ISE) and one intronic splicing silencer element (ISS) in between the two ISE (represented in Fig. 3.5), that are in balance when the mutations are not there,[108] and the occurrence of mutations in positions where the elements occur can impair the splicing pattern.

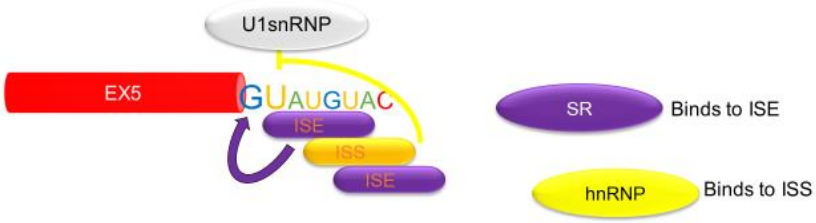


Figure 3.5: Schematic representation of splicing elements located in the 5'ss consensus, that can be disrupted if any mutation occurs. The splicing mutations located in the same splicing consensus disrupt the local bind of splicing elements, such as Intronic splicing silencer (ISS) a local binding of heterogeneous Ribonucleoproteins (hnRNP), that repress the splicing and may cause the exon skipping and Intronic splicing enhancer (ISE), a local binding of serin-arginin rich proteins (SR) wich enhance the splicing and may promote the exon inclusion. Additionally, the local binding of Small Ribonucleoproteins (U1snRNP), which recognise the splice site, to begin the splicing process is in the same sequence.

Protein analysis

The impact of the four splicing mutations was also studied at the protein level in CFBE cells and HEK cells transiently transfected with each minigene. To this end, total protein was harvested 48h post-transfection, and CFTR protein was analysed by WB. Results from CFBE cells transiently expressing each minigene revealed that cells transfected with wt minigenes exhibited two bands, a fully mature glycosylated CFTR band C (170kDa), a core-glycosylated fraction band B (140kDa)[109] and the 127 kDa band A (Fig. 3.6). [110,111]

In both cell lines (CFBE and HEK), the protein resulting from the four splicing mutations presented two lower molecular weight bands, namely: one band between 140 and 170KDa and another lower band of less than 140 kDa (Fig 3.6). These results can suggest that the upper band (between 140 and 170KDa) is a fully-glycosylated form of CFTR protein without 30 aminoacids (correspondent to

90bp of the exon 5) and the other lower band corresponds to a core-glycosylated form of this shorter protein.

However, the lack of exon 5, corresponds to the loss of only 30 aminoacids (30aa), and as the average molecular weight of an aminoacid is 100Da, 30aa will correspond to 3KDa. The 7% polyacrylamide gel used here is not sufficient to show this small difference. But, if these mutations produce a protein with different conformation, the protein can migrate differently in the gel.

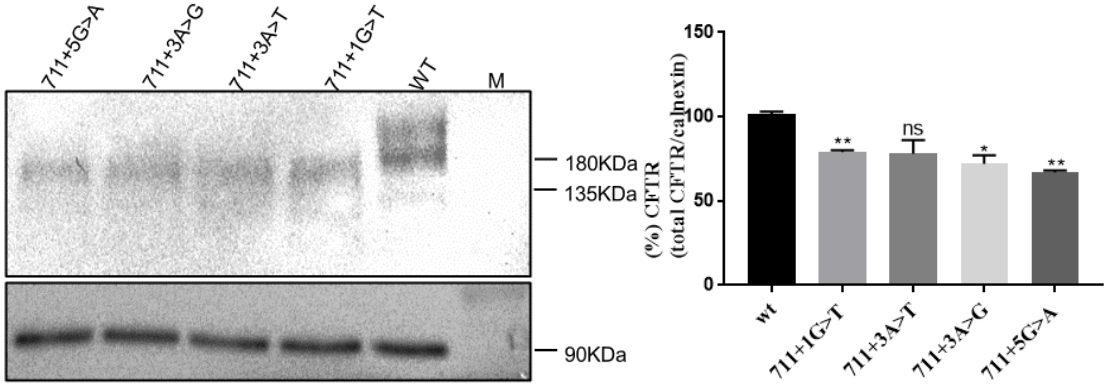


Figure 3.6: Analysis of the protein obtained from CFBE cells transiently transfected with wt, 711+1G>T, 711+3A>T, 711+3A>G and 711+5G>A minigenes. In the left panel, a representative membrane from WB shows the migration pattern of protein obtained from transiently transfected cells, with different molecular weights. In the right panel, the quantification protein of gels (n=3) indicates the quantification of the total amount of CFTR in wt and all mutants. The comparison between wt and each mutant were performed using an unpaired t-test, and the statistical differences are shown as (**) when 0.001<P<0.01, (*) when P=0.02 and (ns) when the data are not statistically different. Data are expressed as mean ± SEM of three independent experiments.

In fact, one study using a construction of CFTR cDNA, with deletion of exon 5, revealed that cells transfected with this Δexon5 CFTR expressed only a core-glycosylated form of CFTR with 140KDa. It also showed that neither incubation at lower temperature (26º), nor stimulation of HEK 293T cells with forskolin cause rescue in glycosylation and processing of the Δexon5 CFTR protein, indicating that this mutation was not processed properly.[109]

The first intracellular loop (aminoacid 139-194) connects transmembrane segments II and III, and 30 of the residues in this loop are encoded by exon 5.[112] It was shown that deleting 30 aminoacids from the first intracellular loop of CFTR affects both processing and function of the CFTR Cl^- channel.[109] So most likely our mutants have a similar defect. To confirm this, further studies should include the functional characterization of the protein to determine the channel activity.

The minigenes are double-tagged to allow to verify the localization of the protein in all four mutants. These minigenes have eGFP in the N-terminus to see the total amount of the protein and flag in the 4th extracellular loop to see only the protein located in the PM. With this strategy was possible to observe that the protein of cells expressing the wt mini-gene is located in the PM, but in all four mutants the eGFP fluorescence is decreased and that the protein is mislocalized in the PM (Fig. 3.7).

These data would corroborate previous results obtained 21 years ago, which showed that the skipping of exon 5 caused misprocessing of the CFTR protein, that this aberrant protein does not reach the membrane and that it is retained primarily in the intracellular membrane.[109]

Altogether these data validated the use of minigene constructs to study the effect of splicing mutations and showed that the mutations produce truncated proteins that may affect either their processing or trafficking.

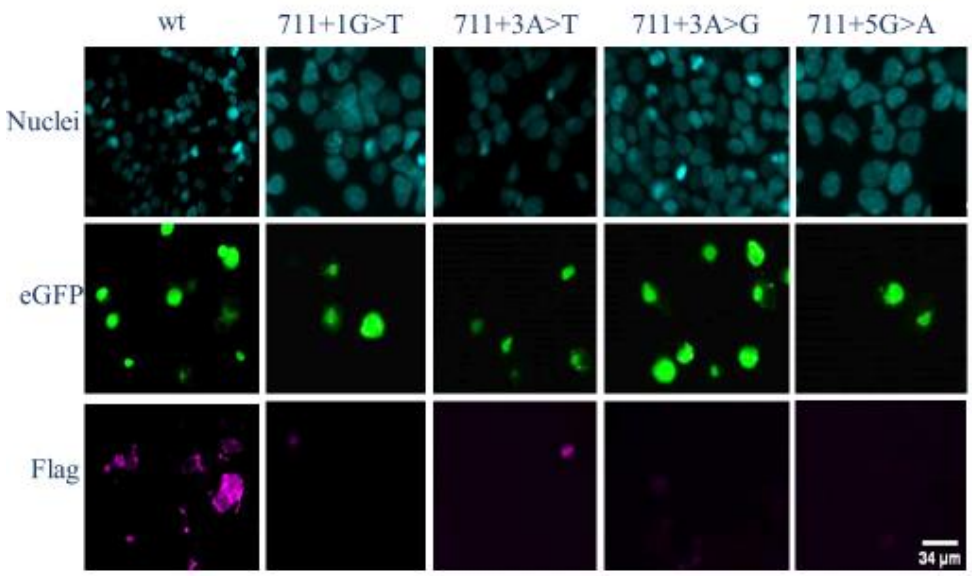


Figure 3.7: Characterization of localization of the CFTR protein of HEK 293T cells expressing the wt and all four mutants minigene by immunofluorescence. The CFTR protein in HEK293T cells expressing three splicing mutations is not localized in the membrane.

3.2.3. Effect of the AON on all splicing mutations

It was shown that the four splicing mutations in the same splicing consensus (711+1G>T, +3A>T, +3A>G and +5G>A) cause full skipping of exon 5, and reduce the amount of correctly spliced transcripts to about 95%.

After studying the impact of four mutations located in the same splicing consensus, we aimed to explore whether we were able to correct the splicing defect caused by the mutations using an AON. Only one AON was designed to cover all the mutations as they are in the same splicing consensus.

For that, an AON was designed 20 bases downstream from the donor splice site (Fig. 3.8), and it was transiently co-transfected with each minigene in HEK 293T cells followed by the analysis of the splicing correction at the mRNA level by semi-quantitative and quantitative PCR. Data from both analyses showed that the same AON could correct all four mutations at the mRNA level, with major correction seen in 711+3A>G (as shown in figure 3.9 and 3.10).

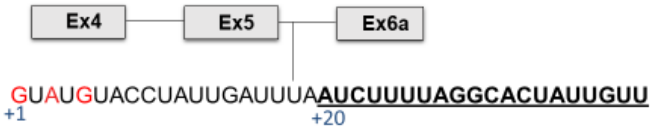


Figure 3.8: Schematic representation of the position of AON in IVS5.

Quantification of transcripts showed that the same AON efficiently corrected the splicing mutations in the range of 10-65%, being efficient the longer the distance from the canonical site, being more efficient to correct the 711+3A>T than 711+3A>G. In each mutation, the AON corrected in the different proportion. Despite being no statistical difference for the mutation 711+1G>T, the AON promoted the exon inclusion from 2% to 12%. For the other mutations it efficiently corrected, namely, for 711+3A>T from 20% t0 50%, for 711+3A>G from 9% to 19% and for 711+5G>A from 4% to 43%

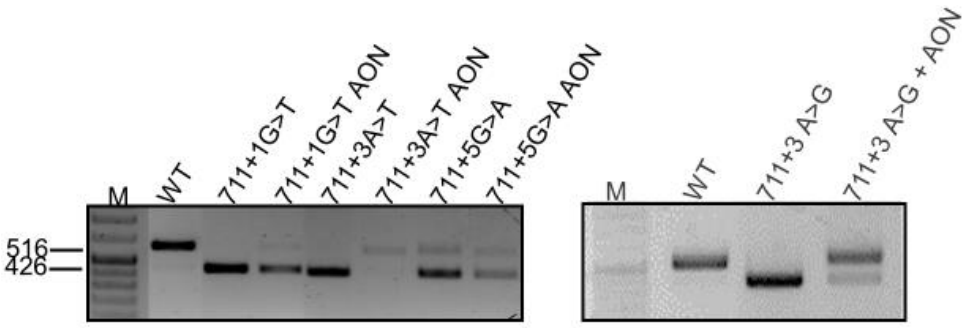


Figure 3.9: Correction of alternative splicing accessed by RT-PCR. Representative gel from the electrophoretic analysis showing the enhancement of inclusion of exon 5 (upper band) in cells co-transfected with AON and each mutant minigene**.**

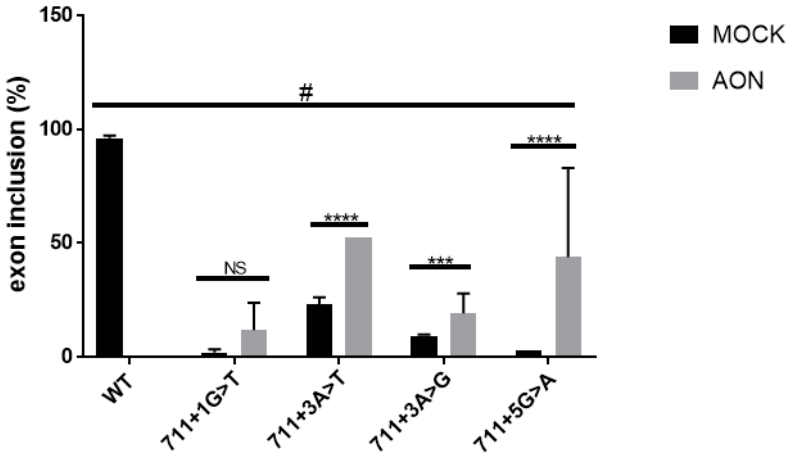
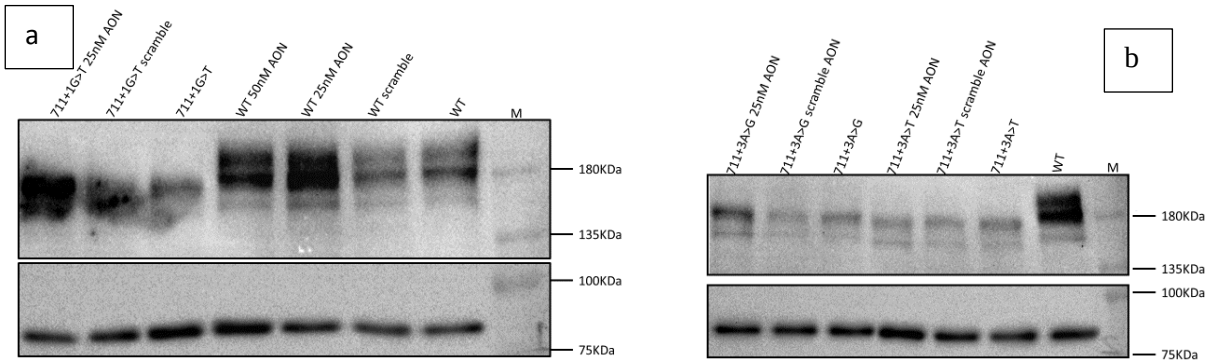


Figure 3.10: Correction of alternative splicing accessed by qPCR. Real-time qPCR showing that the same AON corrects alternative splicing caused by all four mutations in the range of to 10-65% at increased efficiencies the longer the distance to the canonical splice site, and, for position +3. It being more efficient to correct the A>T than A>G : 711+1 G>T < 711+3 A>G < 711+3 A>T < 711+5 G>A. The differences between the mock and treated cells expressing each minigene were accessed by unpaired t-test, and the differences between the wt and all mutants were accessed by 2way ANOVA. The differences were statistically considered when P<0.05. Data are expressed as mean ± SEM of three independent experiments.

To access whether the AON increases the total CFTR protein, we performed WB with lysates from HEK 293T cells co-expressing the AON and each mutant, and it was observed that the protein increased after the AON treatment, however, only in mutant 711+3A>G the differences were statistically different (Fig.3.11). A non-specific AON was used as a negative control (referred here as scramble), and it was observed that this scrambled AON did not have any effect in correcting the mutants.These data suggest that the AON effect is sequence specific for intron 5.



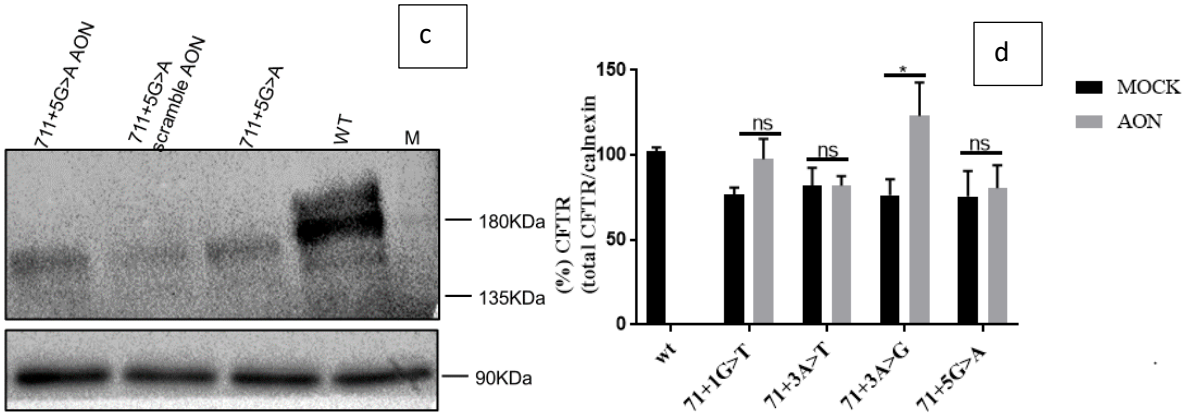
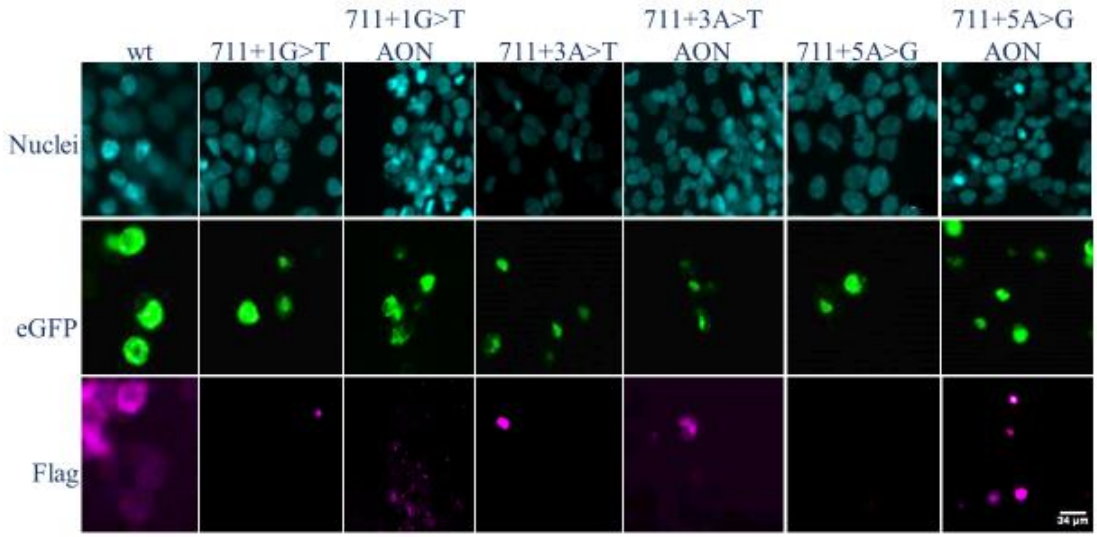


Figure 3.11: Correction of four splicing mutations in the same splicing consensus by AON strategy at the protein level (a-c). Representative western blot analysis of wt and four splicing (711+1G>T, 711+3A>T, 711+3A>G and 711+5G>A) mutations. Cells expressing the wt minigene were submitted to all treatments used in cells expressing the mutants minigenes to evaluate the toxicity and malicious effect of the treatments. (d) Quantification of total protein from WB analysis from two independent experiments showing that the same AON could enhance the total amount of protein of all mutants with exception in mutant 711+3A>T. The differences between each mutant before and after treatment were assessed using t-test and were considered only statistically different when $P<0.05$ (*). Data are expressed as mean ± SEM of two independent experiments.

All four splicing mutations may promote misprocessing of the CFTR protein, leading to a mislocalization of protein at the surface of the cell. To see whether the AON could promote the proper processing and traffic of the CFTR protein to the plasma membrane, HEK 293T cells transiently expressing each minigene treated and not treated with AON were used for immunofluorescence (Fig. 3.12). The AON promoted the expression of the flag in the mutants, which we could not detect in the non-treated cells.

Altogether, these data with the mRNA and protein analyses indicate that the same AON can correct mutations localized in the same splicing consensus site, suggesting that that patients with one of this four mutations could benefit from this approach.



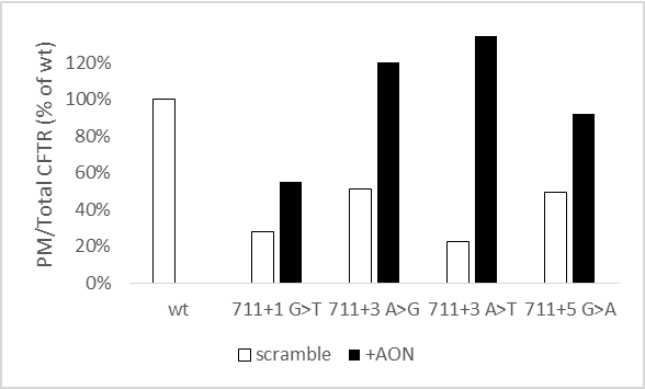


Figure 3.12: Correction of four splicing mutations in the same splicing consensus by AON strategy at the protein level accessed by immunofluorescence. Representative images of AON treated and non-treated cells. Quantification of CFTR at the PM. The CFTR at the PM was normalized to total expression levels (eGFP fluorescence). Data are the mean of three different experiments and shown as mean ± SEM.

Herein, we have hypothesized how all four mutations impair the correct splicing and that the occurrence of these mutations disrupt the fine-tuned balance of factors that regulate splice site selection, such as ISE and ISS elements (Fig. 3.13). Furthermore, it is well known that the U1 small RNA (U1 snRNA), the first component of the spliceosome that guide the recognition of the donor splice site interacts with the first six nucleotides of the intron,[108] thus suggesting that the mutations also impair the base pair of U1snRNA and the donor site.

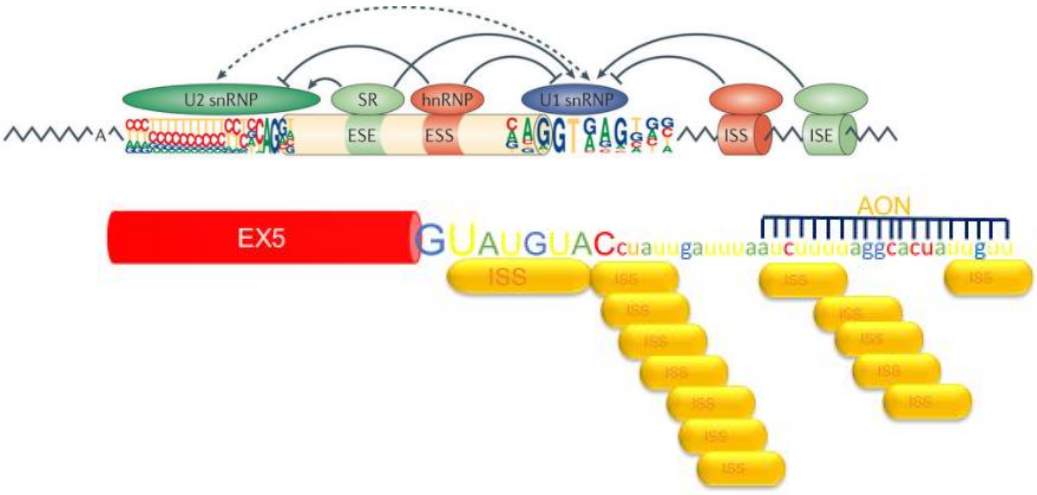


Figure 3.13: Schematic representation of the cis-acting splicing regulatory elements and trans-acting splicing factors. These cis-acting elements can be classified as intronic splicing silencer (ISS), intronic splicing enhancer (ISE), exonic splicing silencer (ESS) and exonic splicing enhancer (ESE) based on their location and activity. In general ISS and ESS recruit splicing factors (heterogeneous nuclear ribonucleoproteins - hnRNPs) that inhibit the recognition of nearby splice sites, while ISE and ESE recruit splicing factors (serine-arginine rich proteins family - SR) that inhibit the recognition of nearby splice sites, thus inhibiting and promoting the splicing, respectively. The dashed arrow represents the first step of the spliceosome assembly and coloured letters represent the consensus motifs of the splice sites. Splicing at a particular site can be activated or repressed when these auxiliary proteins bind nearby enhancer or silencer regions—to recruit or inhibit assembly/stability of spliceosomal components, respectively. downstream of the splice site, in the IVS5 the sequence is rich in ISS's, as it bears 14 inhibitor elements until the position +32. Indeed, the AON designed here totally covers 6 of these ISS and represses the binding of heterogeneous nuclear ribonucleoproteins A1 (hnRNP A1 – which usually promote exon exclusion) (lower scheme). Adapted from[113]

However, this is just a speculative argument which does not explain why exon five is excluded in mutation 711+1G>T, since a novel mutation (in the same position), the 711+1G>A mutation was recently described as leading to a full retention of intron 5.[114]

It was also found that the same AON could correct all these four mutations with different efficiencies, but the question now is, how does it correct them? One explanation may be that downstream of the splice site, the sequence is rich in ISS's, as it bears 14 inhibitor elements until the position +32,[107]. Indeed, the AON designed here totally covers 6 of these ISS, and represses the binding of heterogeneous nuclear ribonucleoproteins A1 (hnRNP A1 – which usually promote exon exclusion)[115].

This suggests that these ISSs within the first 32 nucleotides of IVS5 may play an important role in the exon 5 exclusion. Further studies should be conducted to confirm these hypotheses and find out which elements are promoting the exclusion of the exon 5.

Section IV: **Conclusions and future perspectives**

The first part of this MSc work consisted in studies of the possible involvement of the most common commonly found among Portuguese population in causing non-CF respiratory diseases, and we conclude that F508del was the most common mutation in all individuals together, even among carriers, with the incidence of 62%. The second most common mutation was the 3272-26 A>G mutation extensively described in Portugal, but not included in the CF database.Two mutations (G542X and F508del) and one sequence variation (G576A) were found among non-CF patients with respiratory diseases, in which COPD patients had the highest incidence (4.9%) of mutated alleles and asthma patients had the lowest incidence (0.9%) of the mutated alleles. Furthermore, patients with asthma and COPD presented only F508del mutation.

In this study, only the most common mutations that occur in Portugal were screened, and with this, some important or novel mutations can be missed, and mainly for the non-CF individuals with respiratory diseases, where rare mutations can be found. It would be beneficial to conduct additional CFTR genetic analysis in a larger cohort of mutations to rule out the CFTR as a contributor of other respiratory diseases.

The second part was the characterization of the impact of 711+1G>T, 711+3A>T, 711A>G and 711+5G>A mutations located in the same splice consensus and their correction with an AON approach. For that end, a splicing reporter minigene already constructed in our lab was used, and it was concluded that these mutations lead to total in-frame skipping of exon 5 with a more pronounced decrease in the inclusion of exon 5, the closer the mutation is to the splice site. The skipping of exon 5 leads to the mislocalization of the CFTR protein in the PM. The unique AON constructed for this study was able to correct all mutations, being more efficient in correcting mutations that lie further from the canonical splice site. These data suggest that all patients with one of these four mutations can benefit with this AON and it can help to improve the quality of life of these patients.

Neither the splicing elements involved in skipping of exon 5 nor the mechanism of action of the AON used in this study are not entirely understood, further studies are necessary such as the multiple deletions of intron 5 and 4 to search for the most important splicing elements involved in the skipping of exon 5 and to see how the same AON could promote such results in all four mutations, even in the +1 mutations that are very difficult to correct.

Section V: References

1. Farrell, P. M. The prevalence of cystic fibrosis in the European Union. *J. Cyst. Fibros.* **7,** 450–453 (2008).
2. European Cystic Fibrosis Society. ECFS Patient Registry Annual Data Report. 1–139 (2014).
3. O'Sullivan, B. P. & Freedman, S. D. Cystic Fibrosis. in *SpringerReference* 67–75 (Springer-Verlag, 2008). doi:10.1007/SpringerReference_135807
4. Bobadilla, J. L., Macek, M., Fine, J. P. & Farrell, P. M. Cystic fibrosis: A worldwide analysis of CFTR mutations - Correlation with incidence data and application to screening. *Hum. Mutat.* **19,** 575–606 (2002).
5. Ferec, C. & Cutting, G. R. Assessing the disease-liability of mutations in CFTR. *Cold Spring Harb. Perspect. Med.* **2,** (2012).
6. De Boeck, K. & Amaral, M. D. Progress in therapies for cystic fibrosis. *Lancet Respir. Med.* **4,** 662–674 (2016).
7. O'Sullivan, B. P. & Freedman, S. D. Cystic fibrosis. *Lancet* **373,** 1891–1904 (2009).
8. Amaral, M. D. & Balch, W. E. Hallmarks of therapeutic management of the cystic fibrosis functional landscape. *J. Cyst. Fibros.* **14,** 687–699 (2015).
9. Somayaji, R., Ramos, K. J., Kapnadak, S. G., Aitken, M. L. & Goss, C. H. Common clinical features of CF (respiratory disease and exocrine pancreatic insufficiency). *Presse Med.* **46,** e109–e124 (2017).
10. Cohen-Cymberknoh, M., Shoseyov, D. & Kerem, E. Managing cystic fibrosis: Strategies that increase life expectancy and improve quality of life. *Am. J. Respir. Crit. Care Med.* **183,** 1463–1471 (2011).
11. Tabori, H. *et al.* Abdominal symptoms in cystic fibrosis and their relation to genotype, history, clinical and laboratory findings. *PLoS One* **12,** e0174463 (2017).
12. Amaral, M. D. Novel personalized therapies for cystic fibrosis: Treating the basic defect in all patients. *J. Intern. Med.* **277,** 155–166 (2015).
13. Farrell, P. M. *et al.* Diagnosis of Cystic Fibrosis: Consensus Guidelines from the Cystic Fibrosis Foundation. *J. Pediatr.* **181,** S4–S15.e1 (2017).
14. Farrell, P. M. *et al.* Diagnosis of Cystic Fibrosis in Screened Populations. *J. Pediatr.* **181S,** S33–S44.e2 (2017).
15. Sosnay, P. R. *et al.* Applying Cystic Fibrosis Transmembrane Conductance Regulator Genetics and CFTR2 Data to Facilitate Diagnoses. *J. Pediatr.* **181,** S27–S32 (2017).
16. Farrell, P. M. P. *et al.* Guidelines for Diagnosis of Cystic Fibrosis in Newborns through Older Adults: Cystic Fibrosis Foundation Consensus Report. *J Pediatr* **153,** S4–S14 (2008).
17. Wang, L. & Freedman, S. D. Laboratory tests for the diagnosis of cystic fibrosis. *Am.J.Clin.Pathol.* **117 Suppl:,** 0 (2002).
18. Karczeski, B. a. & Cutting, G. R. Diagnosis of Cystic Fibrosis, CFTR-Related Disease and Screening. **34,** 69–76 (2006).
19. Dalcin, P. D. T. R. & Abreu E Silva, F. A. De. Cystic fibrosis in adults: diagnostic and therapeutic aspects. *J. Bras. Pneumol. publicaçã̃o Of. da Soc. Bras. Pneumol. e Tisilogia* **34,** 107–17 (2008).
20. Yankaskas, J. R., Marshall, B. C., Sufian, B., Simon, R. H. & Rodman, D. Cystic Fibrosis Adult Care: Consensus Conference Report. *Chest* **125,** 1S–39S (2004).
21. Ikpa, P. T., Bijvelds, M. J. C. & de Jonge, H. R. Cystic fibrosis: toward personalized therapies. *Int. J. Biochem. Cell Biol.* **52,** 192–200 (2014).
22. Wilschanski, M. Novel therapeutic approaches for cystic fibrosis. *Discov. Med.* **15,** 127–33 (2013).
23. Elborn, J. S. *et al.* Report of the European Respiratory Society/European Cystic Fibrosis Society task force on the care of adults with cystic fibrosis. *Eur. Respir. J.* **47,** 420–8 (2016).
24. Silva, A., Amorim, A., Azevedo, P., Lopes, C. & Gamboa, F. Cystic fibrosis - characterization of the adult population in Portugal. *Rev. Port. Pneumol.* **22,** 141–145 (2016).
25. Kerem, B. *et al.* Identification of the cystic fibrosis gene: genetic analysis. *Science* **245,** 1073–80 (1989).
26. Riordan, J. R. *et al.* Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science* **245,** 1066–73 (1989).
27. Fajac, I. & Wainwright, C. E. New treatments targeting the basic defects in cystic fibrosis. *Presse Med.* (2017). doi:10.1016/j.lpm.2017.01.024
28. Pranke, I. M. & Sermet-Gaudelus, I. Biosynthesis of cystic fibrosis transmembrane conductance regulator. *Int. J. Biochem. Cell Biol.* **52,** 26–38 (2014).
29. Sweezey, N., Tchepichev, S., Gagnon, S., Fertuck, K. & O'Brodovich, H. Female gender hormones regulate mRNA levels and function of the rat lung epithelial Na channel. *Am. J. Physiol.* **274,** C379-86 (1998).
30. Riordan, J. R. CFTR function and prospects for therapy. *Annu. Rev. Biochem.* **77,** 701–726 (2008).
31. Moran, O. On the structural organization of the intracellular domains of CFTR. *Int. J. Biochem. Cell Biol.* **52,** 7–14 (2014).
32. Liu, F., Zhang, Z., Csanády, L., Gadsby, D. C. & Chen, J. Molecular Structure of the Human CFTR Ion Channel. *Cell* **169,** 85–95.e8 (2017).

33. Sheppard, D. N. & Welsh, M. J. Structure and function of the CFTR chloride channel. *Physiol. Rev.* **79,** S23-45 (1999).
34. Sagel, S. D. & Accurso, F. J. Monitoring inflammation in CF. Cytokines. *Clin.Rev.Allergy Immunol.* **23,** 41–57 (2002).
35. Guillot, L. *et al.* Lung disease modifier genes in cystic fibrosis. *Int. J. Biochem. Cell Biol.* **52,** 83–93 (2014).
36. Than, B. L. N. *et al.* CFTR is a tumor suppressor gene in murine and human intestinal cancer. *Oncogene* **35,** 4179–4187 (2016).
37. Sallenave, J. Phagocytic and signaling innate immune receptors: are they dysregulated in cystic fibrosis in the fight against Pseudomonas aeruginosa? *Int. J. Biochem. Cell Biol.* **52,** 103–107 (2014).
38. Pittman, J. E. & Ferkol, T. W. The Evolution of Cystic Fibrosis Care. *Chest* **148,** 533–542 (2015).
39. Egan, M. E. Genetics of Cystic Fibrosis: Clinical Implications. *Clin. Chest Med.* **37,** 9–16 (2016).
40. Sosnay, P. R. *et al.* Defining the disease liability of variants in the cystic fibrosis transmembrane conductance regulator gene. *Nat. Genet.* **45,** 1160–7 (2013).
41. Cystic fibrosis mutation database. (2011). Available at: http://www.genet.sickkids.on.ca/Home.html. (Accessed: 30th April 2017)
42. Krvavac, A. & Nayak, R. P. New Horizons in Cystic Fibrosis - A Review. **4,** (2016).
43. FDA. FDA Notification to the Cystic Fibrosis Community on Kalydeco (ivacaftor). (2012). Available at: https://www.fda.gov/drugs/drugsafety/ucm316693.htm. (Accessed: 23rd September 2017)
44. FDA *et al.* Center for Drug Evaluation and. *J. Cyst. Fibros.* **10,** 84–94 (2010).
45. Boyle, M. P. *et al.* A CFTR corrector (lumacaftor) and a CFTR potentiator (ivacaftor) for treatment of patients with cystic fibrosis who have a phe508del CFTR mutation: a phase 2 randomised controlled trial. *Lancet. Respir. Med.* **2,** 527–38 (2014).
46. Wainwright, C. E. *et al.* Lumacaftor–Ivacaftor in Patients with Cystic Fibrosis Homozygous for Phe508del *CFTR. N. Engl. J. Med.* **373,** 220–231 (2015).
47. Davies, J. C. *et al.* Safety, pharmacokinetics, and pharmacodynamics of ivacaftor in patients aged 2-5 years with cystic fibrosis and a CFTR gating mutation (KIWI): an open-label, single-arm study. *Lancet. Respir. Med.* **4,** 107–15 (2016).
48. Harman, K., Dobra, R. & Davies, J. C. Disease-modifying drug therapy in cystic fibrosis. *Paediatr. Respir. Rev.* 8–10 (2017). doi:10.1016/j.prrv.2017.03.008
49. Bell, S. C., De Boeck, K. & Amaral, M. D. New pharmacological approaches for cystic fibrosis: Promises, progress, pitfalls. *Pharmacol. Ther.* **145,** 19–34 (2015).
50. Welsh, M. J. & Smith, A. E. Molecular mechanisms of CFTR chloride channel dysfunction in cystic fibrosis. *Cell* **73,** 1251–1254 (1993).
51. Cholon, D. M. *et al.* Potentiator ivacaftor abrogates pharmacological correction of ΔF508 CFTR in cystic fibrosis. *Sci. Transl. Med.* **6,** 246ra96 (2014).
52. Van Goor, F. *et al.* Correction of the F508del-CFTR protein processing defect in vitro by the investigational drug VX-809. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 18843–18848 (2011).
53. Pranke, I. M. *et al.* Correction of CFTR function in nasal epithelial cells from cystic fibrosis patients predicts improvement of respiratory function by CFTR modulators. *Sci. Rep.* **7,** 7375 (2017).
54. Kaufman, M. B. Pharmaceutical Approval Update. *P T* **41,** 476–8 (2016).
55. Trezise, A. E. O. *et al.* Non-PCR methods for the analysis of CFTR transcripts. *J. Cyst. Fibros.* **3,** 25–28 (2004).
56. Balestra, D. *et al.* Regulation of a strong F9 cryptic 5'ss by intrinsic elements and by combination of tailored U1snRNAs with antisense oligonucleotides. *Hum. Mol. Genet.* **24,** 4809–4816 (2015).
57. Igreja, S., Clarke, L. A., Botelho, H. M., Marques, L. & Amaral, M. D. Correction of a Cystic Fibrosis Splicing Mutation by Antisense Oligonucleotides. *Hum. Mutat.* **37,** 209–215 (2016).
58. Friedman, K. J. *et al.* Correction of aberrant splicing of the cystic fibrosis transmembrane conductance regulator (CFTR) gene by antisense oligonucleotides. *J. Biol. Chem.* **274,** 36193–9 (1999).
59. Cystic Fibrosis Foundation. (2017). Available at: https://www.cff.org/News/News-Archive/2017/FDA-Approves-Ivacaftor-for-23-Additional-CFTR-Mutations/. (Accessed: 20th July 2017)
60. Mall, M. A. & Galietta, L. J. V. Targeting ion channels in cystic fibrosis. *J. Cyst. Fibros.* **14,** 561–570 (2015).
61. Dekkers, J. F. *et al.* Potentiator synergy in rectal organoids carrying S1251N, G551D, or F508del CFTR mutations. *J. Cyst. Fibros.* **15,** 568–578 (2016).
62. Awatade, N. T. *et al.* Measurements of Functional Responses in Human Primary Lung Cells as a Basis for Personalized Therapy for Cystic Fibrosis. *EBioMedicine* **2,** 147–153 (2015).
63. Spitali, P. & Aartsma-Rus, A. Splice modulating therapies for human disease. *Cell* **148,** 1085–1088 (2012).
64. Tazi, J., Bakkour, N. & Stamm, S. Alternative splicing and disease. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1792,** 14–26 (2009).

65. Lee, M. *et al.* Systematic Computational Identification of Variants That Activate Exonic and Intronic Cryptic Splice Sites. *Am. J. Hum. Genet.* **100,** 751–765 (2017).
66. Douglas, A. G. L. & Wood, M. J. A. RNA splicing: Disease and therapy. *Brief. Funct. Genomics* **10,** 151–164 (2011).
67. Coelho, A. I. *et al.* Functional correction by antisense therapy of a splicing mutation in the GALT gene. *Eur. J. Hum. Genet.* **23,** 500–506 (2015).
68. Soukarieh, O. *et al.* Exonic Splicing Mutations Are More Prevalent than Currently Estimated and Can Be Predicted by Using In Silico Tools. *PLoS Genet.* **12,** 1–26 (2016).
69. Castellani, C. *et al.* Consensus on the use and interpretation of cystic fibrosis mutation analysis in clinical practice. *J Cyst Fibros* **7,** 179–196 (2008).
70. Hillman, R. T., Green, R. E. & Brenner, S. E. An unappreciated role for RNA surveillance. *Genome Biol.* **5,** R8 (2004).
71. Dujardin, G., Commandeur, D., Le Jossic-Corcos, C., Ferec, C. & Corcos, L. Splicing defects in the CFTR gene: Minigene analysis of two mutations, 1811+1G>C and 1898+3A>G. *J. Cyst. Fibros.* **10,** 212–216 (2011).
72. Aartsma-Rus, A. & van Ommen, G.-J. B. Antisense-mediated exon skipping: a versatile tool with therapeutic and research applications. *RNA* **13,** 1609–24 (2007).
73. Ramalho, A. S. *et al.* Five percent of normal cystic fibrosis transmembrane conductance regulator mRNA ameliorates the severity of pulmonary disease in cystic fibrosis. *Am. J. Respir. Cell Mol. Biol.* **27,** 619–627 (2002).
74. Chu, C. S., Trapnell, B. C., Curristin, S. M., Cutting, G. R. & Crystal, R. G. Extensive posttranscriptional deletion of the coding sequences for part of nucleotide-binding fold 1 in respiratory epithelial mRNA transcripts of the cystic fibrosis transmembrane conductance regulator gene is not associated with the clinical manifestati. *J. Clin. Invest.* **90,** 785–90 (1992).
75. Ferrie, R. M. *et al.* Development, Multiplexing, and Application of ARMS Tests for Common Mutations in the CFTR Gene. *Am. J. Hum. Genet* **51,** 251–262 (1992).
76. Newton, C. R. *et al.* Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res.* **17,** 2503–2516 (1989).
77. Ahlawat, S., Sharma, R., Maitra, A., Roy, M. & Tantia, M. S. Designing, optimization and validation of tetra-primer ARMS PCR protocol for genotyping mutations in caprine Fec genes. *Meta Gene* **2,** 439–449 (2014).
78. Ye, S., Humphries, S. & Green, F. Allele specific amplification by tetra-primer PCR. *Nucleic Acids Res.* **20,** 1152 (1992).
79. Grangeia, A. *et al.* Molecular characterization of the cystic fibrosis transmembrane conductance regulator gene in congenital absence of the vas deferens. doi:10.1097/GIM.0b013e3180318aaf
80. Bombieri, C. *et al.* Complete mutational screening of the CFTR gene in 120 patients with pulmonary disease. *Hum. Genet.* **103,** 718–722 (1998).
81. Pignatti, P. F., Bombieri, C., Marigo, C., Benetazzo, M. & Luisetti, M. Increased incidence of cystic fibrosis gene mutations in adults with disseminated bronchiectasis. *Hum. Mol. Genet.* **4,** 635–639 (1995).
82. Cohn, J. A. *et al.* Relation between mutations of the cystic fibrosis gene and idiopathic pancreatitis. *N. Engl. J. Med.* **339,** 653–8 (1998).
83. Sharer, N. *et al.* Mutations of the cystic fibrosis gene in patients with chronic pancreatitis. *N. Engl. J. Med.* **339,** 645–52 (1998).
84. Akai, S. *et al.* Delta F508 mutation of cystic fibrosis gene is not found in chronic bronchitis with severe obstruction in Japan. *Am. Rev. Respir. Dis.* **146,** 781–3 (1992).
85. Artlich, A. *et al.* Common CFTR mutations are not likely to predispose to chronic bronchitis in northern Germany. *Hum. Genet.* **95,** 226–8 (1995).
86. World Health Organisation. The molecular genetic epidemiology of cystic fibrosis. 26 (2004).
87. CFTR2. Clinical and Functional translation of CFTR. (2011). Available at: https://www.cftr2.org/mutation/general/711%252B5G-%253EA/. (Accessed: 19th September 2017)
88. Bienvenu, T. *et al.* Molecular diagnosis of congenital bilateral absence of the vas deferens: analyses of the CFTR gene in 64 French patients. *Ann. Genet.* **40,** 5–9 (1997).
89. Polizzi, A. *et al.* Genotype-phenotype correlation in cystic fibrosis patients bearing [H939R;H949L] allele. *Genet. Mol. Biol.* **34,** 416–420 (2011).
90. Ziętkiewicz, E. *et al.* CFTR mutations spectrum and the efficiency of molecular diagnostics in Polish cystic fibrosis patients. *PLoS One* **9,** (2014).
91. Silva, A., Amorim, A., Azevedo, P., Lopes, C. & Gamboa, F. Cystic fibrosis - characterization of the adult population in Portugal. *Rev. Port. Pneumol.* (2016). doi:10.1016/j.rppnen.2015.12.010
92. Wong, L. J. C. *et al.* Improved detection of CFTR mutations in Southern California Hispanic CF patients. *Hum. Mutat.* **18,** 296–307 (2001).

93. Dal'Maso, V. B. *et al.* Diagnostic contribution of molecular analysis of the cystic fibrosis transmembrane conductance regulator gene in patients suspected of having mild or atypical cystic fibrosis. *J. Bras. Pneumol.* **39,** 181–189 (2013).
94. Chillón, M. *et al.* Identification of a new missense mutation (P205S) in the first transmembrane domain of the CFTR gene associated with a mild cystic fibrosis phenotype. *Hum. Mol. Genet.* **2,** 1741–1742 (1993).
95. Siryani, I. *et al.* Distribution of cystic fibrosis transmembrane conductance regulator (CFTR) Mutations in a cohort of patients residing in Palestine. *PLoS One* **10,** 1–9 (2015).
96. Mendes, F. *et al.* Unusually common cystic fibrosis mutation in Portugal encodes a misprocessed protein. *Biochem. Biophys. Res. Commun.* **311,** 665–671 (2003).
97. Beck, S. *et al.* Cystic fibrosis patients with the 3272-26A→G mutation have mild disease, leaky alternative mRNA splicing, and CFTR protein at the cell membrane. *Hum. Mutat.* **14,** 133–144 (1999).
98. Tzetis, M. *et al.* CFTR gene mutations - Including three novel nucleotide substitutions - And haplotype background in patients with asthma, disseminated bronchiectasis and chronic obstructive pulmonary disease. *Hum. Genet.* **108,** 216–221 (2001).
99. Raju, S. V., Solomon, G. M., Dransfield, M. T. & Rowe, S. M. Acquired Cystic Fibrosis Transmembrane Conductance Regulator Dysfunction in Chronic Bronchitis and Other Diseases of Mucus Clearance. *Clin. Chest Med.* **37,** 147–158 (2016).
100. Raju, S. *et al.* Impact of heterozygote CFTR Mutations in COPD patients with Chronic Bronchitis. *Respir. Res.* **15,** 18 (2014).
101. Casals, T., Ramos, M. D., Larriba, J. G. S., Nunes, V. & Estivill, X. High heterogeneity for cystic fibrosis in Spanish families: 75 mutations account for 90% of chromosomes. *Hum. Genet.* **101,** 365–370 (1997).
102. Felício, V., Ramalho, A. S., Igreja, S. & Amaral, M. D. mRNA-based detection of rare CFTR mutations improve genetic diagnosis of cystic fibrosis in populations with high genetic heterogeneity. *Clin. Genet.* **91,** 476–481 (2017).
103. Zielenski, J. *et al.* Identification of mutations in exons 1 through 8 of the cystic fibrosis transmembrane conductance regulator (CFTR) gene. *Genomics* **10,** 229–35 (1991).
104. Petreska, L., Kočeva, S., Gordova-muratovska, A., Nestorov, R. & Efremov, G. D. Identification of two new mutations (711 + 3A>G and V1397E) in Cf chromosomes of Albanian and Macedonian origin. *Hum. Mol. Genet.* **3,** 999–1000 (1994).
105. Bisceglia, L. *et al.* Development of RNA-SSCP Protocols for the Identification and Screening of CFTR Mutations : Identification of Two New Mutations. *Hum. Mutat.* **140,** 136–140 (1994).
106. Fonknechten, N., Chomel, J. C., Kitzis, A., Kahn, A. & Kaplan, J. C. Skipping of exon 5 as a consequence of the 711 + 1 g→t mutation in the CFTR gene. *Hum. Mol. Genet.* **1,** 281–282 (1992).
107. FO Desmet, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, B. C. Human Splicing Finder. (2013). Available at: http://www.umd.be/HSF3/4DACTION/AW_QuickMutant. (Accessed: 19th September 2017)
108. Tazi, J., Bakkour, N. & Stamm, S. Alternative splicing and disease. *Biochimica et Biophysica Acta - Molecular Basis of Disease* (2009). doi:10.1016/j.bbadis.2008.09.017
109. Xie, J., Drumm, M. L., Zhao, J., Ma, J. & Davis, P. B. Human epithelial cystic fibrosis transmembrane conductance regulator without exon 5 maintains partial chloride channel function in intracellular membranes. *Biophys. J.* **71,** 3148–3156 (1996).
110. Trou, P. *et al.* Rapid detection of the mature form of cystic fibrosis transmembrane regulator by surface plasmon resonance. *R. Soc. Chem.* **7,** 226–236 (2015).
111. Gregory, R. J. *et al.* Expression and characterization of the cystic fibrosis transmembrane conductance regulator. *Nature* **347,** 382–386 (1990).
112. Xie, J., Drumm, M. L., Ma, J. & Davis, P. B. Intracellular loop between transmembrane segments {IV} and {V} of cystic fibrosis transmembrane conductance regulator is involved in regulation of chloride channel conductance state. *J Biol Chem* **270,** 28084–28091 (1995).
113. Matera, A. G. & Wang, Z. A day in the life of the spliceosome. *Nat. Rev. Mol. Cell Biol.* **15,** 108–121 (2014).
114. Lullo, A. Di *et al.* An ' ex vivo model ' contributing to the diagnosis and evaluation of new drugs in cystic fibrosis. *Acta Otorhinolaryngol. Ital.* 1–7 (2016). doi:10.14639/0392-100X-1328
115. Carmel, I., Tal, S., Vig, I. & Ast, G. Comparative analysis detects dependencies among the 5' splice-site positions. *RNA* **10,** 828–40 (2004).
116. Exon Mine data base. (2016). Available at: http://www.imm.fm.ul.pt/exonmine. (Accessed: 1st January 2016)

Section VI: Appendices

Appendix 1 – Introduction, Material and Methods

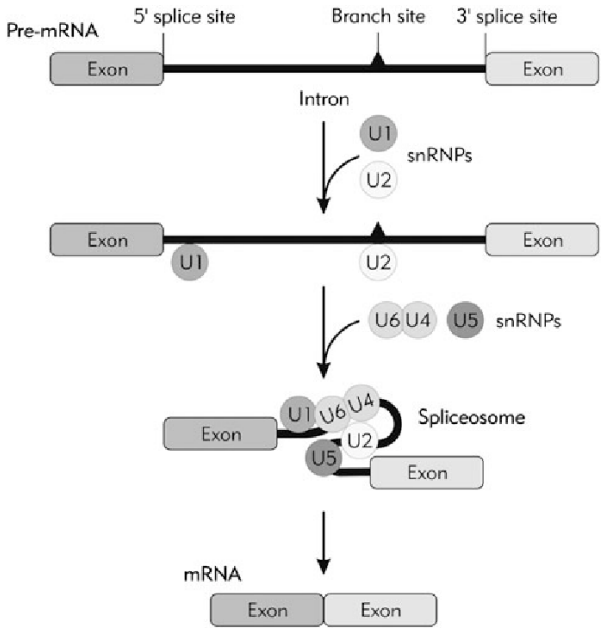


Figure 1S: schematic representation of splicing process. U1 binds to the 5' splice site (5′-ss) by complementarity and U2 binds to the branch point. The triple snRNPs (U4, U5 and U6) complex moves-in to associate with the assembling spliceosome. Then, the U4 leaves the complex, allowing the replacement of U1 by U6 that interacts with U2 to bring the branch point into proximity to the 5′-ss. At this point, the first transesterification reaction cleaves the 5'-ss of the intron of the downstream exon and attaches it to the branch point. U5 then brings the 3'-ss of the upstream exon and 5'-ss of the downstream exon into proximity with each other, allowing a second transesterification reaction that cleaves the 3'-ss of the upstream exon.

1. RNA extraction

Using the NucleoSpin® RNA kit from MACHERY-NAGEL (MN) (Duren, Germany), cells and tissues are lysed by incubation in a solution containing significant amounts of chaotropic ions. This lysis buffer immediately inactivates RNases, which are present in virtually all biological materials, creates an appropriate binding condition which favour adsorption of RNA to the silica membrane. The remaining contaminating DNA that also bound to the silica membrane is removed by a rDnase solution, applied directly onto the membrane during the preparation. Simple washing steps with two different buffers remove salts, metabolites and macromolecular cellular components. Pure RNA is finally eluted under low ionic strength conditions with RNase-free water.

Following extraction, the RNA was quantified using Nanodrop 200 spectrophotometer, and purity was assessed by the A260/280 ratio. Only RNA samples with a ratio above 2.0 were deemed acceptable. The extracted samples were stored at -80ºC.

2. cDNA synthesis

cDNA from the previously extracted RNA was synthesised with NZY M-MuLV REVERSE TRANSCRIPTASE kit. The enzyme from this kit is a recombinant form of the Reverse Transcriptase from the Moloney Murine Leukemia Virus (M-MuLV) purified from *Escherichia coli*. This enzyme lacks 3′→5′ exonuclease activity and has no RNase H activity, enabling improved synthesis of full-

length cDNA even for long mRNA, using random priming. Furthermore, it gives high yields of first strand cDNA up to 7Kb.

1µg of RNA was incubated with 50ng of a Random hexamer, 0.5mM of dNTP Mix and RNase water free up to 17µl for 5min at 65ºC to promote the denaturation of the secondary structure of the RNA facilitating the binding of the oligo-dTs to the poly-A (poly adenosine) strand of mRNAs and enhancing reverse transcription yields. Then the pre-mix containing RNA, random hexamer, dNTP Mix and water was chilled on ice, followed by addition of 2µl of 10x Reaction Buffer (500mM Tris-HCl, pH 8.3, 750mM KCl, 30mM $MgCl_2$ and 100mM DTT), 1µl of Nzy Ribonuclease Inhibitor and 200 units of Nzy M-MuLV Reverse Transcriptase. Afterwards, this mixture was incubated for 10 min a 25ºC, then for 50min at 37ºC and inactivated for 15min at 70ºC.

The resulting cDNA strand was used for reverse transcription- polymerase chain reaction (RT-PCR) or were stored at -20ºC.

3. Transformation of competent bacteria and extraction of plasmids

200µl of competent bacteria made in our lab were incubated with the hydrolysis product (described in 2.2.2) for 30 min on ice, followed by heat-shock for 1.5min at 42ºC, incubation for 2 min in ice, and then incubation in antibiotic-free LB medium for one hour at 37ºC at 220rpm. Bacteria were then centrifuged for 2 min at 6000g. The supernatant was discarded and the pellet was resuspended in the remaining supernatant medium. This suspension was then plated onto LB agar (Sigma-Aldrich) supplemented with 100µg of Ampicillin and incubated overnight at 37ºC.

The recombinant bacteria were then grown in LB medium (Sigma-Aldrich) supplemented with 100µg of ampicillin and incubated overnight at 37ºC, where then the plasmids were extracted.

The plasmids were extracted using NZYMiniprep kit (NZYTech) following the manufacturers' instructions, and the concentration was measured as described in this section (1). Afterwards, to confirm the introduction of each pontual mutation, the extracted plasmids were sent to StabVida for sequencing, with the primers described in Table 5.

4. qPCR: quantitative analysis

mRNA expression was analysed using quantitative real-time RT-PCR (qRT-PCR). cDNA from lung tissue and cells were lysed to assess the mRNA expression of *CFTR* gene.

qPCR was performed in 96 well plates (Bio-Rad Laboratories) with Evagreen reaction mixture (Bio-Rad Laboratories) using Cx96 real-time PCR machine. The primers for the distinction of wt and alternative transcript were requested and selected by Dr Luka Clark using ExonMine database.[116] The primers utilized in qPCR amplify across exon boundaries between exon 4 and exon 5 to minimise potential background amplification of products from genomic DNA.

5 µl of the template cDNA (diluted 1:5), 250 nM of each forward and reverse primer (outlined in this section – Table 3), and 1x Evagreen PCR reaction mixture per well was used to prepare the qPCR reaction. The optimized cycle conditions for Evagreen was used: 10 seconds (sec) at 95 ºC and 30 sec at 60ºC, these steps were repeated 40 cycles. To confirm amplification of specific product melting curve was performed with a temperature gradient from 65 to 95ºC.

Relative abundance of mRNA was measured using samples run in triplicate the Bio-Rad CFX Manager 2.0 software (Bio-Rad, 1845000).

Transcript levels of the *CFTR* gene were calculated by normalizing to an endogenous gene, Adenylate cyclase-associated protein 1 (Cap-1). The mean threshold cycle (CT) of the targets genes was calculated by subtracting the mean CT of the control gene, ΔCT. Then, the mean ΔCT was calculated for a chosen condition, control condition. This mean ΔCT was calculated from several replicates, and it was then subtracted from the individual ΔCT for each assay, giving the ΔΔCT for each gene compared

to the chosen condition. The fold difference in gene expression was calculated by the relative quantification method using the mathematical equation $2^{-\Delta\Delta CT}$

Table 1: Description of all primers used for complete sequencing of CFTR

Name	Orientation	Sequence 5'→3'
T7	Foward	
Ex3F	Foward	GGA TAG AGA GCT GGC TTC
CF5ZR*	Foward	TTG TAC CAG CTC ACT ACC T
Ex5F	Foward	CTC CTT TCC AAC AAC CTG AAC
B3R	Foward	AAT GTA ACA GCC TTC TGG GAG
C1R	Foward	GTG GAG GTC AAC GAG CAA GA
D2R	Foward	GCA AAC TTG ACT GAA CTG GA
E1R	Foward	AGA TTC TCC AAA GAT ATA GC
Ex18F	Foward	AAC TCC AGC ATA GAT GTG G
Ex22F	Foward	AGC AGT TGA TGT GCT TGG C

*mainly used for confirmation of the pontual mutations on IVS5.

Table 2: Description of mutagenic primers for splicing mutations introducing

Name	Orientation	Sequence 5'→3'
711+1G>T.F	Foward	CCT GAA CAA ATT TGA TGA ATT ATG TAC CTA TTG
711+1G>T.R	Reverse	CAATAGGTACATAATTCATCAAATTTGTTCAGG
711+3A>G.F	Foward	CTG AAC AAA TTT GAT GAA GTG TGT ACC TAT TGA TTT AAT C
711+3A>G.R	Reverse	GATTAAATCAATAGGTACACACTTCATCAAATTTGTTCAG
711+3A>T.F	Foward	CTG AAC AAA TTT GAT GAA GTT TGT ACC TAT TGA TTT AAT C
711+3A>T.R	Reverse	GATTAAATCAATAGGTACAAACTTCATCAAATTTGTTCAG
711+5G>A.F	Foward	GAACAAATTTGATGAAGTATATACCTATTGATTTAATC
711+5G>A.R	Reverse	GAT TAA ATC AAT AGG TAT ATA CTT CAT CAA ATT TGT TC

Table 3: Description of primers used in qPCR technique

Name	Orientation	Sequence 5'→3'	Function
Cap-1.F	Foward	ATGCACCGTGGGTATGCAG	Amplify endogenous gene
Cap-1.R	Reverse	AAGCAGCGAGTCAAATGCCT	
Ex5Rev	Reverse	TCAGGTTGTTGGAAAGGAGAC	Amplify only wt transcripts
Ex4/6Rev	Reverse	CCAATGCAAGTCCCTTCTTATAAATC	Amplify only transcripts with no exon 5
Ex4Fwd common	Foward	TGACCCGGATAACAAGGAGG	Common